JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01399.x
Available online at: interscience.wiley.com

PAPER

PHYSICAL ANTHROPOLOGY

Kristen M. Hartnett,[1] *Ph.D.*

Analysis of Age-at-Death Estimation Using Data from a New, Modern Autopsy Sample—Part I: Pubic Bone*,†

ABSTRACT: This research tests the accuracy of age estimation from the pubic bone. Specimens were collected from decedents of known age, sex, and race at the Forensic Science Center (FSC) in Phoenix, Arizona. The collection consists of pubic bones and fourth rib ends from 419 males and 211 females, ranging in age from 18 to 99. Age-at-death was estimated by three observers using the Suchey–Brooks method. The correlation results indicate that there are significant differences in the observed versus actual ages ($r = 0.68169$, $p < 0.001$) and that there are significant inter-observer differences. No significant differences were found in the intra-observer tests. The FSC pubic bones were sorted based on morphology without knowing age. New descriptions and age ranges were created. A phase seven was described and is comprised of males and females over 70 years of age-at-death.

KEYWORDS: forensic science, forensic anthropology, age estimation, pubic symphysis, pubic bone, osteology

While forensic anthropologists use a number of adult skeletal aging techniques, one of the most commonly used standards is for the pubic bone, which is often referred to as the pubic symphysis method in anthropological literature and practice. The pubic bone method relies on visual assessments of the topography and appearance of the surface of pubic symphysis, which demonstrates predictable degenerative patterns corresponding to increasing age. This research evaluates the accuracy of the pubic bone method developed by Suchey and colleagues (1,2). To perform this research, a large sample of pubic bones ($N = 626$) of known age was created to test the accuracy and precision of the accepted standards. In addition, sternal ends of the fourth rib from the same individuals were collected at the same time and will be presented in Part II of this research, which is the topic of another article.

Major criticisms of the pubic bone age-phase method center on sample size and composition, population specificity, high inter- and intra-observer error rates, and the large age ranges of the phases established in the original studies (3–7). In addition, the Suchey–Brooks (SB) pubic bone sample is currently not accessible for further study. Recent statistical re-evaluations of age-phase methods suggest that the classical linear regression and other models used in the original studies may contribute to the reduced accuracy levels because they do not accurately reflect true age phases, but transitions from age phase to age phase in the reference sample (3,4,8–11). Based on the criticisms of the pubic bone techniques in the literature, three goals were established for this research: (i) to create a new, documented sample for future research and education, (ii) to evaluate the current SB standards on a large, modern, and diverse sample, and (iii) to propose revisions that increase the accuracy and precision of the method.

Materials

Skeletal specimens were collected from decedents of known age, sex, and race during examination at the Maricopa County Forensic Science Center (FSC) in Phoenix, Arizona, from January 11, 2005 through June 30, 2006. In addition, specimens were collected from Barrow Neurological Institute in Phoenix, Arizona, when available, to supplement the FSC sample. The total collection ($N = 630$) consists of pubic bones and bilateral fourth rib ends from 419 males and 211 females, ranging in age from 18 to 99 (Fig. 1). Individuals classified by the medicolegal system at the FSC as Asian ($n = 4$), Black ($n = 20$), Caucasian (including Hispanics; $n = 598$), and Native American ($n = 8$) were represented in the sample. In addition to the demographic information on age, sex, and race, information regarding the drug and alcohol history was obtained when available. Summary statistics such as counts, means, medians, modes, standard deviations, and ranges were calculated to help describe the different aspects of the research sample (Table 1). Because of preservation and preparation problems, as well as pathological or traumatic conditions in some individuals, both pubic bones and fourth rib ends were obtained and analyzed from 582 individuals (387 males and 195 females). Twenty-eight males and 14 females lack scoreable rib ends, yielding a total sample of 584 individuals for the rib analysis. Four males and two females are

[1]Office of Chief Medical Examiner, 520 First Avenue, New York, NY 10016.

*This research was funded by a Wenner-Gren Foundation Dissertation Fieldwork Grant, #7284, the ASU Dean's Dissertation Writing Fellowship, the ASU Division of Graduate Studies Completion Fellowship, the Ellis R. Kerley Forensic Sciences Foundation Scholarship, an ASU GPSA Research Grant, a Forensic Sciences Foundation Acorn Grant, and an ASU Dept. of Anthropology Research Grant.

†Presented at the 59th Annual Meeting of the American Academy of Forensic Sciences, February 22, 2007, in San Antonio, TX.

Received 24 Nov. 2008; and in revised form 17 June 2009; accepted 27 June 2009.

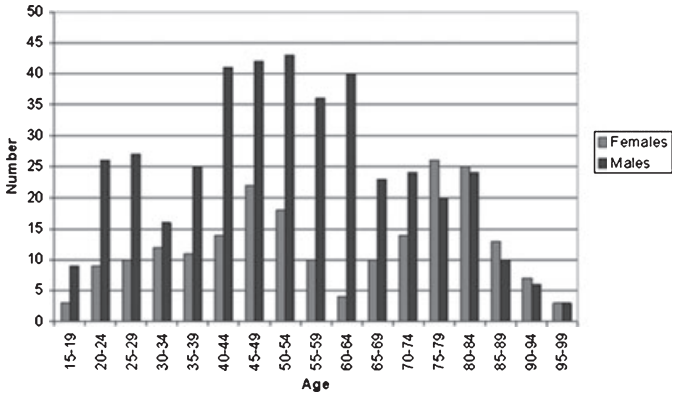


FIG. 1—*Age and sex distribution of the total sample.*

represented only by rib ends, yielding a total sample of 620 individuals for the pubic bone analysis.

Methods

When a decedent arrived at the Maricopa County FSC and fit the study criteria, the legal next of kin of the deceased person was contacted to request consent to include the decedent in the study. The phone conversation was conducted in compliance with the protocol previously submitted to and approved by the Arizona State University Human Subjects' Institutional Review Board and was recorded to ensure proper consent was obtained. After each specimen was collected, a copy of the phone transcript and a letter with the contact information of the researcher was mailed to the next of kin for their records.

Once permission to use the individual in the study was granted from the legal next of kin, two fourth rib segments and the pubic bones were removed with an oscillating saw or large clipper tool during examination. Each rib was cut at approximately 2 inches from the sternal end, and the pubic bones were extracted with two cuts on each side through the superior and inferior rami. Soft tissue was manipulated with a scalpel and forceps to facilitate extraction of the bone segments. The bone segments were macerated and then dried on racks in the laboratory. Each set was labeled with a random code number and the sex of the individual, according to the protocols set forth in the approved Institutional Review Board submission.

Age was estimated using the SB casts and written phase descriptions for the pubic bone (2). Two volunteers, in addition to the author, participated in an inter- and intra-observer error study. One volunteer was a forensic anthropologist with a Ph.D. and over 20 years of experience in the field, whereas the second volunteer was a physical anthropologist with an M.A. and approximately 4 years of experience in the field. Each volunteer estimated the age of 50 males and 50 females using the SB pubic bone casts. After several days, each observer was asked to rescore a random selection of 15 males and 15 females that the author culled from the main sample to assess intra-observer reliability. All skeletal analyses in this research were performed with the knowledge of the sex of the individual, but without any prior knowledge of age.

TABLE 1—*Basic statistics describing the ages of individuals in the FSC collection by sex.*

	Count	Mean	Median	Mode	St. Dev.	Range
FSC males	409	52.2	52.0	48.0	21.50	18–97
FSC females	197	57.9	54.5	75.0	21.45	18–99
BNI males	10	70.7	73.0	–	20.0	42–93
BNI females	14	75.4	74.0	75.0	19.97	58–93
Total males	419	52.6	52.0	52.0	19.0	18–97
Total females	211	59.2	58.0	75.0	21.4	18–99
Sample total	630	54.8	53.0	48.0	20.1	18–99

FSC, Forensic Science Center.

Next, the pubic bones were segregated by sex and sorted and seriated separately based on observed morphological characteristics of the symphyseal face as well as bone quality. Two observers jointly inspected both symphyseal faces of each individual and placed them in broad groups with similar morphology. Initially, the elements were placed into the broad and general categories of young, middle, and old adult based on visual inspection of the features of the symphyseal face (i.e., rim formation, appearance of face, bone quality). After this initial sort into broad groups, both observers further subdivided the groups using more specific character states and additional morphological characteristics. Together, both observers inspected every individual in each group to confirm that they fit the morphology of the group. Written descriptions of the general characteristics for each group were created, the actual ages were recorded, and the mean, standard deviation, and range were calculated.

For the pubic bone, the morphological features that were considered to be the most discriminatory by the two observers were the extent of the ridge and furrow system on the face, the degree of symphyseal rim formation/deterioration, and bone quality. An individual was placed into the young group if he/she had extensive ridges and furrows on the face, an unformed or incomplete rim, and good bone quality. A person was placed into the middle-aged group if the pubic bones had slight ridges and furrows or remnants on the face, a partial or complete rim, and good bone quality. Finally, an individual was included in the older adult group if there were no ridges and furrows on the face, the rim was complete, and the bone quality was fair to poor.

Several diverse statistical tests were run on the data collected from the bone segments, ranging from simple summary statistics to complicated Bayesian analyses. The summary statistics, correlation, and linear regression procedures are discussed later. The Bayesian transition analysis results will be discussed in a future study. All statistical analyses were performed using the SAS Statistical Software version 9.1.2. Descriptive statistics were calculated for the different types of data collected, and Spearman's coefficient of rank correlation was used to examine the relationship between the observed age-phase estimates and the actual phases for the three observers. In addition, Spearman's correlation was used to test intra-observer reliability.

For comparison with the SB aging method, least-squares linear regression analyses with resulting r values were conducted on this new sample. Generally, traditional regression analyses are inappropriate for the categorical data produced in this study because regression is typically used when the data are continuous. In addition, regression has been shown to underage the elderly, and overage young individuals (11) and is very sensitive to the age composition of the reference sample (3). Strictly for ease of comparison with previous studies, however, regression analyses were performed.

Results

The results of the correlation analyses describing the relationship between the known age-at-death and SB phase (actual) and the estimated phase (observed) are presented in Table 2. The correlation

TABLE 2—*Correlation of estimated versus actual Suchey–Brooks phases for each observer in the combined sex sample.*

	Count	Spearman's *r*	*p* Value
Author	620	0.64439	<0.0001
Observer 1	100	0.69936	<0.0001
Observer 2	100	0.58048	<0.0001

TABLE 3—*Degree of association of estimated pubic bone scores between three observers in the combined sex sample.*

	Count	Spearman's *r*	*p* Value
Author/Observer 1	100	0.79356	<0.0001
Author/Observer 2	100	0.62142	<0.0001
Obs. 1/Obs. 2	100	0.69210	<0.0001

values for all three observers are positive, highly statistically significant at $p < 0.0001$, and fairly strong, ranging from approximately 0.58 to 0.70. Table 3 compares the difference between the estimates of each observer (inter-observer error) for the entire sample. *R* values for all three observers in relation to each other were positive, moderate (ranging from 0.62 to 0.79), and very significantly correlated ($p < 0.0001$). The sample was then segregated into males and females, and these comparisons were performed again. The correlation values were higher for females than for males (males 0.53–0.61; females 0.66–0.77).

Each of the three observers re-examined 15 males and 15 females. Males and females were also analyzed separately to determine whether either sex held more consistent intra-observer values. The correlation value for the three observers approach is 0.90, which is a strong and significant correlation between the initial and second estimates (Table 4). All three observers were fairly consistent with themselves in phase estimation using the SB method.

For comparison with previous age estimation studies, least-squares linear regression analyses with resulting *r* values were conducted on this new sample. Phase was run as a continuous, scalar, *X* variable, with actual age as the *Y*, or dependent variable. Regression of the actual numerical age on the SB phase (one through six) was performed for males and females separately. All tests were highly significant ($p < 0.0001$) with strong *r* values (Table 5). The *r* value for the males ($r = 0.867$) was only slightly less than the *r* value for females ($r = 0.877$).

The pubic bones were sorted into seven distinct groups based on shared morphological features. The ages were recorded for all the individuals included in each group, and summary statistics were calculated. Table 6 lists the descriptive statistics for the male study sample when compared to the SB male sample. Phases 1 through 3 are very similar in both samples, with means differing by only a year, but begin to diverge with phase 4. Phases 4 and 5 in the FSC sample have means that are 7–8 years older than in the SB sample. Phase 6 has comparable means.

Table 7 compares the mean ages and ranges of each phase for the female FSC sample, the SB sample (12), and the Balkan and William Bass Donated (WBD) female samples utilized by Berg (13). As with the males, the means for phases 1 through 3 are very similar. The mean age in phase 4 of the FSC sample is older than the mean ages for the SB and Berg phase 4. Phase 5 has fairly similar means across the four samples, but phase 6 is quite variable in the four samples, with the FSC sample having the oldest mean, followed by the WBD sample, SB, and then the Balkan sample. Phase 7 in the FSC sample has a much higher mean age-at-death than the WBD or Balkan sample. In fact, phase 6 in the FSC

TABLE 4—*Correlation of Suchey–Brooks phase estimate 1 versus estimate 2 for each observer in the combined sex, male, and female intra-observer samples.*

	Sample	Count	Spearman's *r*	*p* value
Author	Combined	30	0.89690	<0.0001
Author	Male	15	0.84122	<0.0001
Author	Female	15	0.93691	<0.0001
Observer 1	Combined	30	0.85740	<0.0001
Observer 1	Male	15	0.82330	<0.0001
Observer 1	Female	15	0.91122	<0.0001
Observer 2	Combined	30	0.88978	<0.0001
Observer 2	Male	15	0.90370	<0.0001
Observer 2	Female	15	0.77479	<0.0001

TABLE 5—*Least-squares regression analyses for the male and female pubic bones.*

Sample	df	Sum of Squares	Mean Square	*F*	*p*	*r*	R^2
Males							
Model	1	112492.86	112492.86	1230.90	<0.0001	0.867	0.751
Error	409	37378.90	91.39				
Females							
Model	1	73417.04	73417.04	696.98	<0.0001	0.877	0.770
Error	208	21909.92	105.34				

TABLE 6—*Descriptive statistics for each pubic bone phase in males for the study sample and Suchey–Brooks reference sample.*

	FSC				Suchey-Brooks*			
Phase	*n*	Mean	SD	Range†	*n*	Mean	SD	Range‡
1	14	19.29	1.93	18–22	105	18.5	2.1	15–23
2	14	22.14	1.86	20–26	75	23.4	3.6	19–34
3	36	29.53	6.63	21–44	51	28.7	6.5	21–46
4	69	42.54	8.8	27–61	171	35.2	9.4	23–57
5	90	53.87	8.42	37–72	134	45.6	10.4	27–66
6	34	63.76	8.06	51–83	203	61.2	12.2	34–86
7	96	77	9.33	58–97	–	–	–	–

FSC, Forensic Science Center.
*Data from Brooks and Suchey (2).
†100% of individuals.
‡95% confidence interval.

sample is more comparable to phase 7 in the WBD and Balkan skeletal sample. This pattern may be in part because of the large number of very old individuals in the FSC sample. In all, the discrepancies in mean ages and ranges for the phases are attributed to differences in the criteria used for phase definition as well as differences in the reference sample size and composition.

For the most part, the FSC phase descriptions are modifications of those published by Brooks and Suchey (2). Minor changes were made to the wording of the phase descriptions. In addition, the age ranges and means per phase were adjusted to reflect the phase composition in the FSC sample. The morphological features that were considered to be the best age predictors by the author and second observer were listed first (see Appendix). The descriptions created for phases 1 and 2 in the FSC collection are similar to the Brooks and Suchey (2) phase descriptions and are the same for both males and females, suggesting that the pattern of aging on the pubic bone is similar in both sexes into the early twenties. Beginning around the mid-twenties, the aging patterns diverge slightly. Males and females generally demonstrate similar morphology for phase 3, but some males may exhibit a ventral epiphysis variant in

TABLE 7—*Descriptive statistics for each pubic bone age phase in females.*

	FSC				SB*				BK†				WBD†			
Phase	n	mean	SD	range‡	n	mean	SD	range§	n	mean	SD	range§	n	mean	SD	range§
1	5	19.8	1.33	18–22	48	19.4	2.6	15–24	–	–	–	–	–	–	–	–
2	5	23.2	2.38	20–25	47	25	4.9	19–40	–	–	–	–	–	–	–	–
3	25	31.44	5.12	24–44	44	30.7	8.1	21–53	–	–	–	–	–	–	–	–
4	35	43.26	6.12	33–58	39	38.2	10.9	26–70	4	33.5	8.1	–	6	35.5	3.8	–
5	32	51.47	3.94	44–60	44	48.1	14.6	25–83	12	52.5	12.7	–	18	49.7	5.8	–
6	35	72.34	7.36	56–86	51	60	12.4	42–87	15	56.0	14.1	–	27	64.2	9.0	–
7	56	82.54	7.41	62–99	–	–	–	–	29	74.4	10.4	–	50	74.2	10.9	–

FSC, Forensic Science Center; SB, Suchey–Brooks; BK, Balkan sample; WBD, William Bass Donated Collection.
*Data from Brooks and Suchey (2).
†Data from Berg (13).
‡100% of individuals.
§95% confidence interval.

the formation of the ventral rampart (Fig. 2). According to Suchey (14), males showing this lateral line are usually <35 years old and are often in their twenties. This feature was described in this sample as an epiphysis-type variant on the ventral body occurring only in young males under the age 33. Eight male individuals exhibited the ventral epiphysis variant, with a mean age of 25.6 years, and ranging in age from 22 to 33 years of age-at-death. Individuals in phase 3 have a V-shaped lower extremity that may extend halfway up the ventral face, but is longer on the dorsal face. Some buildup of bone may occur in the space between the upper and lower extremities. The bone quality at this stage is still very good, but the ridges and furrows are worn down and porosity commences.

Once the upper and lower extremities and ventral rampart form, the face has a complete oval outline. A hiatus, or gap, can occur in the upper ventral rim, where the rim has not completed. The ventral hiatus is a characteristic feature of phase 4 in both the SB and FSC phase systems. A ventral hiatus, however, does not persist in all individuals. Changes also can occur on the dorsal rim. Dorsal lipping tends to occur with older individuals and varies in its degree of formation. With dorsal lipping, the dorsal margin appears to have a flare to the edge, when viewed from the dorsal aspect.

In the FSC morphological sort, males and females in phases 4 and 5 are still fairly similar. Beginning with phase 5, however, the bones of the females start to become a little bit lighter and more porous than in males. In addition, the ventral arc often becomes elaborate in females beginning in phase 4. For the most part, both males and females only have remnants, if any, of the ridges and furrows. The rim may be complete or have a ventral hiatus, and the bone shows evidence of deterioration. In females with parturition pits, dorsal lipping tends to be more prominent at this stage. In phase 5, the face becomes more irregular and the rim becomes more erratic in both males and females.

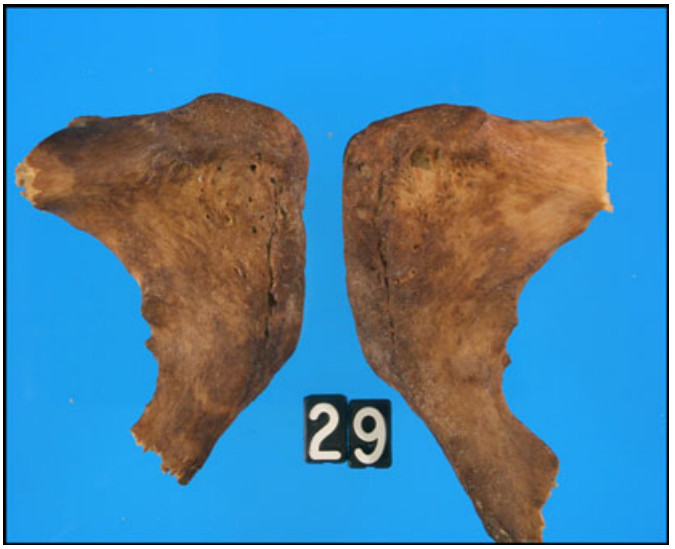


FIG. 2—*Ventral epiphysis variant in male pubic symphyses.*

Distinctive features in older individuals center on the deterioration of the bone. Changes in bone quality occur in both males and females with increasing age. Bone mineral density declines, there is a reduction in stiffness and strength, bone becomes more brittle, and fractures more easily (15). In addition to the changes in bone mineral density, many older individuals also experience a decrease in physical activity and changes in diet that affect the composition of bone. All the biomechanical, hormonal, and cellular changes associated with aging result in a general reduction of bone quality with advancing age.

Many of the following features associated with older individuals involve the "feel" of the bone, which can only be ascertained by picking up the bone to assess relative "weight" and by palpating the surfaces of the bone to determine its texture. Bone weight is a major deciding factor between age phases; often a set of pubic bones will have a younger morphology than what the weight of the bone suggests, but the feature of bone weight alone can be used to move a set of specimens up or down a phase. Bone weight is not affected by the size of the person or the sex of a person, but by age. With increasing age, the bone becomes lighter. The decrease in bone density with increasing age tends to be more dramatic in females than males.

The texture of the bone changes with age as well. In younger individuals, the bone is smooth and dense to the touch. With increasing age, the bone surfaces become rougher and more like sandpaper to the touch (overprocessing of the remains during maceration, however, can also damage specimens so that bone weight and porous surfaces mimic those of old individuals). The dorsal body surface is especially useful for detecting these subtle changes in bone texture. On the ventral surface, small bony projections form along the body and the margin of the obturator foramen. The bone here is usually smoother and more uniform in younger individuals and, in older individuals, the projections are longer and the surface is more irregular to the touch. In addition, the texture and quality of the face yields information about age. In older individuals, the face no longer retains the ridges and furrows, but feels more porous, like sandpaper. Plus, the face may have several depressions or may be completely depressed in relation to the rim.

The rim can undergo erosion and disfigurement, causing the overall shape of the rim to transform from an oval to a very irregular shape. In addition, the rim tends to break down most often at the superior and ventral margin—the same location as the feature characteristic of phase 4, the ventral hiatus. When the rim is breaking down, the edges are more porous and irregular than when a hiatus with rounded edges is present. This feature is very difficult to assess, making it hard to decide between phase 4 and 5 on some occasions (12). The breakdown of the superior/ventral rim may be the precursor to the crenulation feature in some males as discussed later.

In phases 6 and 7, the bone of both males and females loses some integrity. The face loses its oval shape, the bone is light in weight, porosity increases on the face, and the texture of the bony surfaces turns coarse like sandpaper (Fig. 3). Phases 6 and 7 are very similar, so the major deciding factor between 6 and 7 is bone weight. Females tend to have lighter bones at these stages than males, but bone weight in general is diminished, regardless of sex. The lighter bone in females is hypothesized to be attributed to bone deterioration through the aging process but also to osteoporosis and osteopenia (13). According to Resnick and Niwayama (16), bone loss begins at approximately age 35 for females and age 40 for males. Bone loss occurs more rapidly in females than males, in a 4:1 ratio until 80 years of age when the males catch up to the females. In addition, pregnancy and childbearing may have an effect on overall bone health (12), but the extent to which they affect bone loss has not been quantified.

In some females from this study sample, the ventral arc became enlarged and more irregular with increasing age. The etiology of this change is unknown, but may be attributed to the partial ossification of the ligament in older females because of parturition or locomotor stresses. The ventral arc is not present in most males to begin with, and no irregular ligamentous outgrowths in the area of the ventral arc were noted in the males in this sample. One feature in the males that was not observed in the older females was the presence of what Suchey and Katz describe as crenulations. According to Suchey and Katz (12), crenulations are "lacy edgings" surrounding the symphyseal face occurring on some individuals over age 50. Their written description does not do this feature justice; their Figure 11 on page 220 (12) better depicts this feature. The crenulations occur at the superior and ventral aspect of the face and appear as a lytic or sclerotic hiatus that extends toward the pubic tubercle and sometimes underneath the ventral rim. This feature should not be confused with a ventral hiatus in the younger individuals. It can be distinguished from a true ventral hiatus by the overall bone quality of the specimen, and the bone quality at the crenulation. The crenulations in older individuals have very irregular edges and appear more like diseased bone. In younger individuals, the hiatus occurs a little bit lower on the ventral edge instead of more on the superior rim edge, and the edges of the true ventral hiatus are more rounded. In the study sample, 15 male individuals had this crenulation/hiatus variant (Fig. 4). The mean age for this group was 69.8 and ranged in age from 60 to 88.

Discussion and Conclusions

The statistical analysis of the data set suggests that the SB pubic bone aging method is not an extremely accurate method of determining skeletal age-at-death. In general, the correlation values are higher for females than for males, suggesting that there was less interobserver error between observers for females than for males. In addition, this may suggest that females undergo a more regular and predictable pattern of aging in the pubic bone than males, which contradicts much of the known literature (17–20). When the entire sample was segregated by sex and evaluated by the author, there were few differences based on sex.

The statistical analysis of the association between the initial and second phase estimates demonstrated a high degree of intra-observer reliability. Other researchers, however, have found a large amount of intra-observer error using this method. For instance, Saunders et al. (21) reported a disagreement rate between first and second trials of 34.2% for males and 56% for females.

The r values produced in the regression analyses in this study are also higher than r values generated in other regression analyses (13,22). Katz and Suchey (23) published r values of 0.83 for White males, 0.78 for Black males, and 0.79 for Mexican males. Because of the small numbers of non-Caucasian individuals in this study sample (n = 32), regression analyses by race were not conducted. The higher r values created in this study may reflect the composition of the sample; there are a large number of elderly individuals but a small number of young individuals.

While these equations may appear to be a simple way to estimate age based on phase, regression equations are rarely ever used because regression models have been shown to reflect the structure of the reference sample from which the equations were created and tends to underage the elderly and overage the young because cases

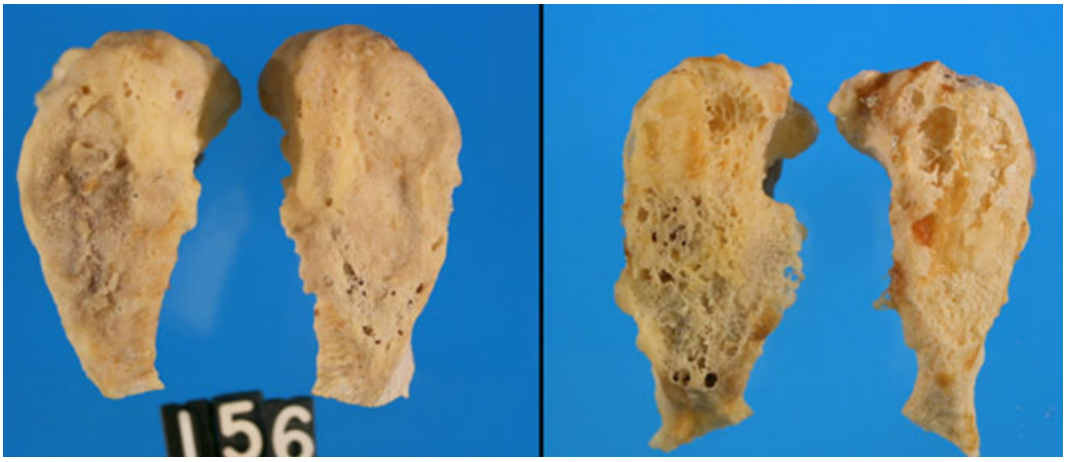


FIG. 3—*Comparison of phase 6 on left and phase 7 on right.*

FIG. 4—*Crenulation variant on older adult males.*

on the extreme ends are pushed toward the mean age of the group (11). Thus, the presence of many older true ages in the reference sample will pull the right side of the regression line up, increasing the slope, and lowering the predicted ages in the younger groups. Plus, because of the poor ability of regression to predict age in older groups, anthropologists have generally accepted that pubic bone morphology cannot be used to determine the age of older individuals accurately. In fact, in one of the original publications, Suchey et al. (1) suggest that age 40 is a reasonable cutoff point under which age can be estimated accurately and above which age cannot be estimated accurately. To test this, they eliminated all individuals older than age 40 in the sample and performed regression analyses on the remaining 405 cases. As they expected, the *r* values and results were better than when older individuals were kept in the analysis.

The FSC phase descriptions and age ranges based on the sample collected at the Maricopa County FSC in Phoenix, AZ, are modifications of those published by Brooks and Suchey (2). Minor changes were made to the wording of the phase descriptions, and age ranges and means per phase were adjusted. Perhaps the most important deviation from the original SB phase descriptions is the description of a phase 7 in both males and females. Many researchers in physical anthropology have proposed that it may be nearly impossible to determine age in elderly skeletons (1,24). Recently, Komar (25) discovered that only 20% of the individuals over 50 years of age were aged correctly in a Bosnian population. Also, in the creation of the SB method, the authors removed individuals greater than 40 years old to try to increase the accuracy of their technique (1). Thus, broad age categories such as 40+ or 50+ have been employed to skirt this issue. This study, as well as that conducted by Berg (13), show that with the creation of a later phase, phase 7, individuals can be placed into later morphological phases and more accurately aged. With the creation of phase 7, one can avoid using age categories like 50+ and instead assign a phase for specimens well into the seventh, eighth, and even ninth decades of life.

In all, the morphological sort and seriation of the pubic bones in the FSC collection indicated that the variation among individuals of the same age and same age phase was much greater than anticipated. The original proposal for this research aimed to reduce the size of the age ranges associated with the age phases. After the sort, the age ranges associated with specific morphological sets of features were, in fact, still fairly large. The original age ranges for stages of pubic bones are wide, but they are wide for the purpose of taking into account this large amount of normal human variation.

Classic criticisms of age-phase methods argue that the age phases reflect the composition of the reference sample from which the phases were created (3,8,26,27). When age phases are constructed based on morphological sorts and seriation of skeletal elements, as described in the results earlier, the reference sample bias is impossible to eliminate with the simple summary statistics used. Thus, the new age-phase categories created in this research based on the morphological sorts of pubic bones are, in fact, reflecting the composition of the reference sample. These revised age-phase categories may be applied to populations in the United States, but should be tested and/or used with caution elsewhere in the world.

This research is of paramount importance because it has generated a modern, autopsy-based, forensic sample of ribs and pubic bones extracted from the same individual for which age, sex, and race are documented. This new sample of specimens, available for all anthropologists to study, will provide a means for independent testing and re-evaluation of rib and pubic bone techniques. The collection of skeletal elements will remain permanently available at the FSC in Phoenix, Arizona.

Acknowledgments

Many thanks to the members of my committee, Dr. Brenda Baker (chair), Dr. Laura Fulginiti, Dr. Robert Williams, and Dr. Charles Merbs, for guiding me through this dissertation research. Many other individuals assisted me in this dissertation work, including all of the staff and doctors at the FSC. Additional specimens were obtained with help from the doctors and the staff of Barrow Neurological Institute and International Institute for the Advancement of Medicine.

References

1. Suchey JM, Wisely DV, Katz D. Evaluation of the Todd and McKern-Stewart methods for aging the male os pubis. In: Reichs KJ, editor. Forensic osteology: advances in the identification of human remains. Springfield, IL: C.C. Thomas, 1986;33–67.
2. Brooks ST, Suchey JM. Skeletal age determination based on the os pubis: a comparison of the Acsádi-Nemeskéri and Suchey-Brooks methods. J Hum Evol 1990;5:227–38.
3. Bocquet-Appel JP, Masset C. Farewell to paleodemography. J Hum Evol 1982;12:321–33.
4. Konigsberg LW, Frankenberg SR, Walker RB. Regress what on what: paleodemographic age estimation as a calibration problem. In: Paine RR, editor. Integrating archaeological demography: multidisciplinary approaches to prehistoric population. Carbondale, IL: Southern Illinois University, 1994;64–88.
5. Jackes M. Building the bases for paleodemographic analysis: adult age determination. In: Katzenberg MA, Saunders SR, editors. Biological anthropology of the human skeleton. New York, NY: Wiley-Liss, Inc, 2000;417–66.
6. Kimmerle EH, Konigsberg LW, Jantz RL, Baraybar JP. Age-at-death estimation through the use of pubic symphyseal data. J Forensic Sci 2008;53:558–68.
7. Djuric M, Djonic D, Nikolic S, Popovic D, Marinkovic J. Evaluation of the Suchey-Brooks method for aging skeletons in the Balkans. J Forensic Sci 2008;52:21–3.
8. Boldsen JL, Milner GR, Konigsberg LW, Wood JW. Transition analysis: a new method for estimating age from skeletons. In: Hoppa RD, Vaupel JW, editors. Paleodemography: age distributions from skeletal samples. United Kingdom: Cambridge University Press, 2002;73–106.
9. Lucy D, Aykroyd RG, Pollard A, Solheim M. A Bayesian approach to adult human age estimation from dental observations by Johanson's age changes. J Forensic Sci 1996;41(2):189–94.
10. Konigsberg LW, Frankenberg SR. Deconstructing death in paleodemography. Am J Phys Anthropol 2002;117:297–309.

11. Aykroyd RG, Lucy D, Pollard AM, Solheim T. Technical note: regression analysis in adult age estimation. Am J Phys Anthropol 1997;104:259–65.
12. Suchey JM, Katz D. Applications of pubic age determination in a forensic setting. In: Reichs KJ, editor. Forensic osteology: advances in the identification of human remains. Springfield, IL: C.C. Thomas, 1998;204–36.
13. Berg GE. Pubic bone age estimation in adult women. J Forensic Sci 2008;53:569–77.
14. Suchey JM. Use of the Suchey-Brooks system for aging the male os pubis. Am J Phys Anthropol 1987;72:259.
15. Martin B. Aging and strength of bone as a structural material. Calcif Tissue Int 1993;1:S34–40.
16. Resnick D, Niwayama G. Diagnosis of bone and joint disorders. Philadelphia, PA: W.B. Saunders Company, 1988.
17. Klepinger LL, Katz D, Micozzi MS, Carroll L. Evaluation of cast methods for estimating age from the os pubis. J Forensic Sci 1992;37:763–70.
18. Gilbert BM. Misapplication to females of the standard for aging the male os pubis. Am J Phys Anthropol 1973;38:39–40.
19. Gilbert BM, McKern TW. A method for aging the female os pubis. Am J Phys Anthropol 1973;38:31–8.
20. Suchey JM. Problems in the aging of females using the os pubis. Am J Phys Anthropol 1979;51:467–70.
21. Saunders SR, Fitzgerald CM, Rogers TL, Dudar C, McKillop H. A test of several methods of skeletal age estimation using a documented archaeological sample. Can Soc Forensic Sci 1992;25:97–118.
22. Katz D, Suchey JM. Age determination of the male os pubis. Am J Phys Anthropol 1986;69:427–35.
23. Katz D, Suchey JM. Race differences in pubic symphyseal aging patterns in the male. Am J Phys Anthropol 1989;80:167–72.
24. Meindl RS, Lovejoy CO. Age changes in the pelvis: implications for paleodemography. In: İşcan MY, editor. Age markers in the human skeleton. Springfield, IL: C.C. Thomas, 1989;137–68.
25. Komar D. Lessons from Srebrenica: the contributions and limitations of physical anthropology in identifying the victims of war crimes. J Forensic Sci 2003;48:713–6.
26. Jackes M. Pubic symphysis age distributions. Am J Phys Anthropol 1985;68:281–99.
27. Bello SM, Thomann A, Signoli M, Dutour O, Andrews P. Age and sex bias in the reconstruction of past population structures. Am J Phys Anthropol 2006;129:24–38.

Additional information and reprint requests:
Kristen M. Hartnett, Ph.D.
Office of Chief Medical Examiner
520 First Avenue
New York, NY 10016
E-mail: khartnett@ocme.nyc.gov

Appendix

Revised pubic bone phase descriptions for the FSC collection

TABLE A1—*Revised pubic bone phase descriptions.*

Phase 1	A clear ridge and furrow system extends from the pubic tubercle onto the inferior ramus. Ridges and furrows are deep and well-defined and do not look worn down. There is no dorsal lipping. Bone is of excellent quality and is firm, heavy, dense, and smooth on the ventral and dorsal body. There is no rim formation. The dorsal plateau is not formed. The ridges and furrows extend to the dorsal edge
Phase 2	The rim is in the process of forming, but mainly consists of a flattening of the ridges on the dorsal aspect of the face and ossific nodules present along the ventral border. Ridges and furrows are still present. The ridges and furrows may appear worn down or flattened, especially on the dorsal aspect of the face. The furrows are becoming shallow. The upper and lower rim edges are not formed. There is no dorsal lipping. The bone quality is very good and the bone is firm, heavy, dense, and smooth on the ventral and dorsal body, with little porosity. The pubic tubercle may appear separate from the face
Phase 3	The lower rim is complete on the dorsal side of the face, and is complete until it ends approximately halfway up the ventral face leaving a medium to fairly large gap between the lower and upper extremities on the ventral face. This enlarged ''V'' is longer on the dorsal side than the ventral side. Some ridges and shallow furrows are still visible, but appear worn down. In some cases, the face is becoming slightly porous. The rim is forming both on the dorsal aspect of the face and the upper and lower extremities. In some cases, there is a rounded buildup of bone in the gap between the upper and lower extremities above the enlarged ''V.'' Bone quality is good; the bone is firm, heavy, dense, and has little porosity. The dorsal surface of the body is smooth, and there are small bony projections near the medial aspect of the obturator foramen. The ventral aspect of the body is not elaborate. Very slight to no dorsal lipping. Quality of bone and rim completion are important deciding factors. Variant: In some cases, a deep line or epiphysis is visible on the ventral aspect parallel to and adjacent to the face (males only)
Phase 4	In most cases, the rim is complete at this stage, but may have a small ventral hiatus on the superior and ventral aspect of the rim. The face is flattened and not depressed. Remnants of ridges and furrows may be visible on the face, especially on the lower half. The quality of bone is good, but the face is beginning to appear more porous. The dorsal and ventral surfaces of the body are roughened and becoming coarse. There is slight dorsal lipping. In females with parturition pits, dorsal lipping can be more pronounced. The ventral arc may be large and elaborate in females
Phase 5	The face is becoming more porous and is depressed, but maintains an oval shape. The face is not irregularly-shaped or erratic. The rim is complete at this stage. In general, the rim is not irregular. Ridges and furrows are absent on the face. There may be some breakdown of the rim on the ventral border, which appears as irregular bone (not rounded/solid). The ventral surface of the body is roughened and irregular, with some bony excrescences. The dorsal surface of the body is coarse and irregular. Projections are present on the medial aspect of the obturator foramen. Bone quality is good to fair; it is losing density and is not smooth. The bone is moderately light in weight. In females the ventral arc is prominent
Phase 6	The face is losing its oval shape and is becoming irregular. The rim is complete, but breaking down, especially on the ventral border. The rim and face are irregular, porous, and macroporous. Bone quality is fair, and the bone is lighter and more porous, even with bony buildup on the ventral body surface. The rim is eroding. The dorsal surface of the bone is rough and coarse. There are no ridges and furrows. Dorsal lipping is present. Projections are present at the medial aspect of the obturator foramen. Bone weight is a major deciding factor between phases 6 and 7
Phase 7	The face and rim are very irregular in shape and are losing integrity. The rim is complete but is eroding and breaking down, especially on the ventral border. There are no ridges and furrows. The face is porous and macroporous. Dorsal lipping is pronounced. Bone quality is poor, and the bone is very light and brittle. Bone weight is an important deciding factor. The dorsal surface of the bone is roughened. The ventral surface of the body is roughened and elaborate. Projections are present at the medial wall of the obturator foramen. The pubic tubercle is elaborate and proliferative. Bone weight is a major deciding factor between phases 6 and 7
VARIANT	VARIANT: The rim is complete except for a lytic/sclerotic appearing hiatus at the superior ventral margin that extends toward the pubic tubercle and sometimes underneath the ventral rim, which should not be confused with a hiatus

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01415.x
Available online at: interscience.wiley.com

PAPER

PHYSICAL ANTHROPOLOGY

Kristen M. Hartnett,[1] *Ph.D.*

Analysis of Age-at-Death Estimation Using Data from a New, Modern Autopsy Sample—Part II: Sternal End of the Fourth Rib*,†

ABSTRACT: This research tests the accuracy of age-at-death estimation from the sternal end of the fourth rib. Age was estimated using the İşcan and Loth casts and written descriptions. The correlation results indicate that there are significant differences in the observed versus actual ages ($r = 0.75329$, $p < 0.001$) and that there are significant interobserver differences. Intraobserver tests showed that no significant differences were found within observers. Results of the rib end analysis compared to the results from the pubic symphyses suggest that the rib performs better than the pubic symphysis in age estimation. The rib ends were sorted based on morphology without prior knowledge of age. Summary statistics were calculated for each new phase, and descriptions were created. A variant form of the rib end was described, and the previously understated feature of bone quality was emphasized.

KEYWORDS: forensic science, forensic anthropology, osteology, age estimation, sternal end of the fourth rib, autopsy sample

This research evaluates the accuracy of the fourth rib method developed by İşcan et al. (1–3) on a large sample of sternal rib ends that was created for this research. Major criticisms of this age-phase method center on sample size and composition, population specificity, moderate to high inter- and intraobserver error rates, and the large age ranges of the phases established in the original studies (4–6). Recent statistical re-evaluations of age-phase methods suggest that statistical models used in the original studies may contribute to the reduced accuracy levels because the models reflect the age structure of the reference sample rather than true age phases (4–10). Based on the criticisms of the fourth rib and other aging techniques in the literature, three goals were established for this research: (i) to create a new, documented sample for future research and education, (ii) to evaluate the current İşcan and Loth (IL) standards on a large, modern, and diverse sample, and (iii) to propose revisions that increase the accuracy and precision of the method.

Materials

Skeletal specimens were collected from decedents of known age, sex, and race during examination at the Maricopa County Forensic Science Center (FSC) in Phoenix, Arizona, from January 11, 2005, through June 30, 2006. In addition, specimens were collected from Barrow Neurological Institute in Phoenix, Arizona, when available, to supplement the FSC sample. The total collection ($N = 630$) consists of pubic bones and bilateral fourth rib ends from 419 males and 211 females, ranging in age from 18 to 99 years (Fig. 1) (see Part I for a full description of the collection).

Methods

When a decedent arrived at the Maricopa County Forensic Science Center and fit the study criteria, the legal next of kin of the deceased person was contacted to request consent to include the decedent in the study. After permission to use the individual in the study was granted, two fourth rib segments and the pubic bones were removed with an oscillating saw or large clipper tool during examination (see Part I for a full description of the sample collection methodology).

Age was estimated using the İşcan and Loth (11) casts and written phase descriptions for the rib ends. The author and two volunteers participated in an inter- and intraobserver error study. Each volunteer estimated the ages of 50 males and 50 females using the İşcan and Loth rib end casts. After several days, each observer was asked to re-score a random selection of 15 males and 15 females the author culled from the main sample to assess intraobserver reliability. All skeletal analyses in this research were performed with the knowledge of the sex of the individual, but without any prior knowledge of age.

Next, the rib ends were segregated by sex and sorted and seriated separately based on observed morphological characteristics. The morphological features that were considered to be the most discriminatory by the two observers were: the depth of the pit, the regularity of the rim edges, and the bone quality. A rib was placed

[1]Office of Chief Medical Examiner, 520 First Avenue, New York, 10016 NY.

*Presented at the 59th Annual Meeting of the American Academy of Forensic Sciences, February 22, 2007, in San Antonio, TX.

†This research was funded by a Wenner-Gren Foundation Dissertation Fieldwork Grant, #7284, the ASU Dean's Dissertation Writing Fellowship, the ASU Division of Graduate Studies Completion Fellowship, the Ellis R. Kerley Forensic Sciences Foundation Scholarship, an ASU GPSA Research Grant, a Forensic Sciences Foundation Acorn Grant, and an ASU Dept. of Anthropology Research Grant.

Received 11 May 2009; and in revised form 7 Aug. 2009; accepted 8 Aug. 2009.

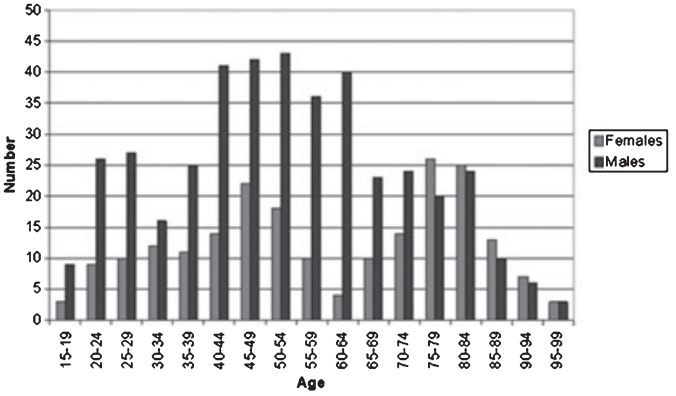


FIG. 1—*Age and sex distribution of the total sample.*

TABLE 1—*Correlation of estimated versus actual İşcan and Loth phases for each observer in the combined sex rib sample.*

	Count	Spearman's *r*	*p*-Value
Author	584	0.69948	<0.0001
Observer 1	100	0.65517	<0.0001
Observer 2	100	0.72027	<0.0001

into the young group if it had a shallow pit, rounded and firm rim edges, and good quality bone. A rib was placed into the middle-aged group if the pit was deep and U-shaped, the rim edges were irregular but still firm and without long bony projections, and the bone quality was good to fair. Finally, the rib was included in the older adult group if the pit was deep and wide, the edges were irregular and thin with long bony projections, and the bone quality was poor.

Several diverse statistical tests were run on the data collected from the bone segments, ranging from simple summary statistics to complicated Bayesian analyses. The summary statistics, correlation, and linear regression procedures are discussed below. The Bayesian transition analysis results will be discussed in a future paper. All statistical analyses were performed using the SAS Statistical Software version 9.1.2. Descriptive statistics were calculated for the different types of data collected, and Spearman's coefficient of rank correlation was used to examine the relationship between the observed age-phase estimates and the actual phases for the three observers. In addition, Spearman's correlation was used to test intraobserver reliability.

Results

The results of the correlation analyses describing the association between the İşcan and Loth phase that each rib end fell into based on known age-at-death (actual) and the İşcan and Loth phase that the rib end was estimated to be (observed) are presented in Table 1. The correlation values for all three observers are positive, highly statistically significant at $p < 0.0001$, and moderate, ranging from 0.66 to 0.72. Table 2 compares the difference between the estimates of each observer (interobserver error) for the rib sample. *R* values for all three observers in relation to each other were positive, moderate (ranging from 0.64 to 0.81), and very significantly correlated ($p < 0.0001$).

Each of the three observers re-examined 15 males and 15 females using the İşcan and Loth phases. Males and females were analyzed separately to determine whether either sex held more

TABLE 2—*Degree of association of estimated rib scores between three observers in the combined sex sample.*

	Count	Spearman's *r*	*p*-Value
Author/Observer 1	100	0.81225	<0.0001
Author/Observer 2	100	0.77845	<0.0001
Observer 1/Observer 2	100	0.64248	<0.0001

TABLE 3—*Correlation of İşcan and Loth phase estimate 1 versus estimate 2 for each observer in the combined sex, male, and female intraobserver samples.*

	Sample	Count	Spearman's *r*	*p*-Value
Author	Combined	30	0.92219	<0.0001
Author	Male	15	0.90100	<0.0001
Author	Female	15	0.95543	<0.0001
Observer 1	Combined	30	0.92183	<0.0001
Observer 1	Male	15	0.92451	<0.0001
Observer 1	Female	15	0.88224	<0.0001
Observer 2	Combined	30	0.95210	<0.0001
Observer 2	Male	15	0.90581	<0.0001
Observer 2	Female	15	1.00000	<0.0001

TABLE 4—*Least squares regression analyses for the male and female rib ends.*

Sample	d.f.	Sum of Squares	Mean Square	F	*p*	*r*	R^2
Males							
Model	1	124,722.19	124,722.19	2705.39	<0.0001	0.935	0.875
Error	385	17,749.01	46.10				
Females							
Model	1	82,251.67	82,251.67	2357.72	<0.0001	0.961	0.924
Error	195	6802.79	34.89				

consistent intraobserver values. The correlation values for the three observers for the rib sample all approach 0.95 (Table 3). All three observers were fairly consistent with themselves in phase estimation using the İşcan and Loth casts and descriptions. If the *r* values for the intraobserver analyses of the rib ends are compared to the *r* values of the pubic symphyses from Part I of this research ($r = 0.90$), it becomes apparent that all three observers were more consistent with themselves (had higher correlation values) with the ribs.

For comparison with previous age estimation studies, least squares linear regression analyses with resulting *r* values were conducted. Phase was run as a continuous X variable, with actual age as the Y, or dependent variable. Regression of the actual numerical age on the İşcan and Loth rib phase (1 through 8) was performed for males and females separately. All tests were highly significant ($p < 0.0001$) with very strong *r* values (Table 4). The *r* value for the males ($r = 0.935$) was less than the *r* value for females ($r = 0.961$), but both were very high and higher than the values for pubic symphysis from Part I of this research (males $r = 0.867$; females $r = 0.877$).

The ribs were sorted into seven distinct categories based on morphological features. Table 5 lists the descriptive statistics for the male FSC sample when compared to the İşcan and Loth male sample (12). Because only individuals 18 years or older were collected for this study, there is no equivalent to the İşcan and Loth phase 0 in males and females, nor phase 1 in females. When phase 1 in the study sample and phase 2 of İşcan and Loth's sample are compared, the means are very similar. Beginning with the

FSC phase 3, the means diverge, and the means for the FSC sample are older. This pattern is likely owing to the fact that this study sample is skewed toward older individuals. Furthermore, in the FSC older phases, the standard deviations are lower, and the ranges are tighter than in the IL phases. For example, in FSC phase 4, the standard deviation is 2.98, and the range is 36–48 years. In the equivalent IL phase (phase 5), the standard deviation is 7.0, and the range is 28–52 years. These differences add up to approximately 4 years for the standard deviation and 12 years for the range. The more homogeneous FSC phases could possibly be attributed to the addition of bone quality and density in the FSC phase descriptions, as well as to the larger sample sizes.

Table 6 illustrates the descriptive statistics for each rib phase for the female FSC sample and the İşcan and Loth sample (12). As with the males, the means for the corresponding early phases are similar, but older in the FSC sample. Beginning with phase 3, the means become more distinct and then approach each other again around the FSC phases 5 and 6 and corresponding İşcan and Loth phases of 6 and 7. The mean of the oldest phase is larger in the study sample than in the original sample. This pattern may be in part owing to the larger number of very old individuals in the FSC sample. The discrepancies in mean ages, ranges, and standard deviations for the phases may be owing to differences in the criteria (i.e., bone quality) used for phase definition, as well as differences in the composition of the reference sample. For example, in the FSC phase 6, the standard deviation is 3.41, and the range is 60–73 years. In the equivalent IL phase (phase 7), the standard deviation is 11.24, and the range is 48–83 years. The differences add up to almost 8 years for the standard deviation and over 20 years for the range.

In general, the FSC category descriptions are modifications of the original İşcan and Loth descriptions. Changes were made to the wording and age ranges, and means per category were adjusted. The features that the observers considered to be the best age predictors were flagged and listed first in the descriptions (see Appendix A). Perhaps the most important alterations were the addition of a variant phase in the males and the recognition of the very important and previously understated role bone quality and density play in phase estimation.

Many of the following features associated with older individuals involve the "feel" of the bone, which can only be ascertained by picking up the bone to assess relative "weight" and by palpating the surfaces of the bone to determine its texture. Bone weight is a major deciding factor between age phases; often, a set of ribs will have a younger morphology than what the weight of the bone suggests, but the feature of bone weight alone can be used to move a set of specimens up or down a phase. Loth (13) noted that, in the Spitalfields collection, there was an apparent contradiction between bone mass and morphology in the rib ends. She cited her own notes as stating, "rib looks phase 8, but feels younger." When this was the case, she stated that her experience has shown that the firmness of the bone is significant and justified assigning a younger phase despite older morphology. Conversely, bone morphology that appears young, but has very poor bone quality, justifies assigning an older phase. Unfortunately, this explanation of bone quality is not described in the original publications and is not always applied. The texture of the bone changes with age as well. In younger individuals, the bone is smooth and dense to the touch. With increasing age, the bone surfaces become rougher and more like sandpaper to the touch. In some instances, the bone of older individuals can become thin and translucent (Fig. 2).

The seven categories described for the males and females in the study sample are modifications of the original phase descriptions created by İşcan et al. (1–3). In the revised descriptions, the author attempted to avoid using character states that involved descriptives, such as "becoming," and relative comparisons to weight and firmness, such as "lighter than," because an observer would likely not have representative real known bones from each phase to compare with the specimen. The revised descriptions also put much more emphasis on bone quality. Here, an individual may be moved up

TABLE 5—*Descriptive statistics for each rib phase in males for the study sample and İşcan and Loth reference sample.*

	FSC				İşcan–Loth*			
Phase	*n*	Mean	SD	Range†	*n*	Mean	SD	Range‡
1	20	20.00	1.45	18–22	4	17.3	0.5	17–18
2	27	24.63	2.00	21–28	25	21.9	2.13	18–25
3	27	32.27	3.69	27–37	27	25.9	3.5	19–33
4	47	42.43	2.98	36–48	12	28.2	3.83	22–35
5	76	52.05	3.50	45–59	14	38.8	7.0	28–52
6	61	63.13	3.53	57–70	17	50.0	11.17	32–71
7	75	80.91	6.60	70–97	17	59.2	9.52	44–85
8	–	–	–	–	12	71.5	10.27	44–85

FSC, Forensic Science Center.
*Data from İşcan et al. (1).
†100% of individuals.
‡95% confidence interval.

TABLE 6—*Descriptive statistics for each rib phase in females for the study sample and İşcan and Loth reference sample.*

	FSC				İşcan–Loth*			
Phase	*n*	Mean	SD	Range†	*n*	Mean	SD	Range‡
1	7	19.57	1.67	18–22	1	14.0	–	–
2	7	25.14	1.17	24–27	5	17.4	1.52	16–20
3	22	32.95	3.17	27–38	5	22.6	1.67	20–24
4	21	43.52	3.08	39–49	10	27.7	4.62	24–40
5	32	51.69	3.31	47–58	17	40.0	12.22	29–77
6	18	67.17	3.41	60–73	18	50.7	14.93	32–79
7	71	81.20	6.95	65–99	16	65.2	11.24	48–83
8	–	–	–	–	11	76.4	8.83	62–90

FSC, Forensic Science Center.
*Data from İşcan et al. (3).
†100% of individuals.
‡95% confidence interval.

FIG. 2—*Extremely thin and nearly translucent rib ends in an older adult female.*

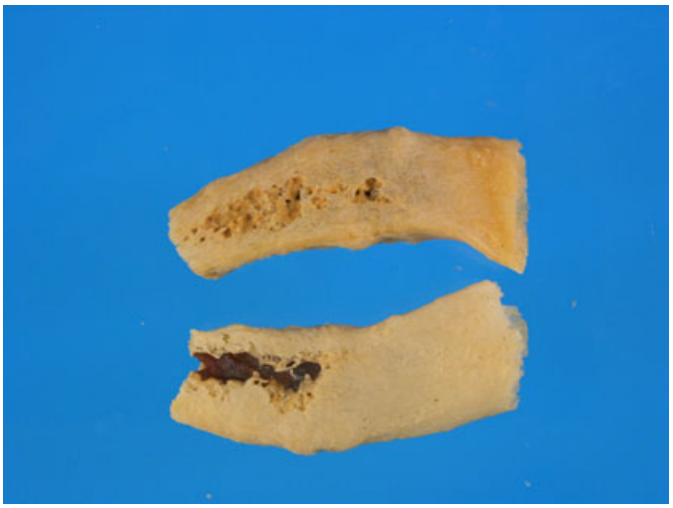


FIG. 3—*Variant form of bony rib extensions in males.*

or down a phase based solely on bone quality. A good example of this occurs with the variant rib form in males that was observed in the study sample (Fig. 3). In some cases, the cartilage had almost completely ossified, with some window formation. These ribs would be classified in the İşcan and Loth system as a phase 8. The ossification, however, is a solid extension that is very good quality bone. The rib body is also very dense, heavy, smooth, and of generally good quality. In these instances, bone quality should be the determining factor. A total of 22 male individuals with this trait were discovered in the study sample, with a mean age of approximately 49 years, and ranging in age from 27 to 69 years. Clearly, the majority of these individuals do not fit in phase 7 or 8 of the İşcan and Loth system. Thus, the very good bone quality should clue in the investigator that some other factor, such as disease or trauma, is at play in the aging process of an individual with this morphology. This variant is too common to ignore ($N = 22$) as simply a case of extreme variation in a few individuals.

Discussion and Conclusions

The statistical analysis of the rib data set suggests that the İşcan and Loth rib end aging method is not an extremely accurate method of determining skeletal age-at-death. While the correlation for all three observers was somewhat strong and significant, it was not as robust as originally hoped, suggesting that the İşcan and Loth system does work but that it could be improved. In general, the correlation values are higher for females than for males, suggesting that there was less interobserver error between observers for females than for males. The moderately high interobserver r values for the rib indicate that the İşcan and Loth method is fairly reliable across observers but could be improved. Supporting this claim is Taylor's (14) research demonstrating poor agreement between estimated phase and actual phase in the ribs. In her study, only 49 of the 155 cases were assigned the real age phase. Other researchers, however, concluded that the rib method was very accurate. For example, İşcan and Loth (12) reported that the overall accuracy was very high, the estimates of the judges fell within one phase of the ideal, and there was minimal disparity between the estimates of judges with different experience levels. Russell et al. (15) concluded that the changes in the morphology of the sternal end of the fourth rib could be used to predict age-at-death with minimal inaccuracy. Perhaps some of these differences could be explained by the fact that the İşcan and Loth phase descriptions leave considerable room for interpretation when it comes to bone weight and quality. This study suggests bone weight and quality play a bigger role in phase assignment, and the revisions to the phases reflect this.

In this study, the correlation r values for the rib ends (0.66–0.72) were higher for all three observers than for the pubic symphysis values obtained in Part I of this research (0.58–0.70), suggesting that the rib method was slightly more accurate than the pubic symphysis method for all three observers. Fulginiti et al. (16,17) also found that the ribs were a more accurate indicator of age than the pubic symphysis and that interobserver error was lower because the ribs showed less variation at any given age. Loth and İşcan (18,19) demonstrated that ribs of both sexes were a more accurate age estimator because ribs were correctly assigned to the ideal chronological phase in a 2:1 ratio to pubes. The combining of adjacent phases, especially in the rib system, may increase interobserver reliability. The drawback, however, is that the associated age ranges would also increase, making the method less useful in narrowing the true age-at-death. Alternatively, weighting certain morphological features reported for each phase could increase the accuracy of the method. By defining the characteristic features of each phase, more emphasis could be placed on those specific features when trying to decide between phases.

There is a strong correlation between initial and second rib estimates in this study; r values for the three observers all approach 0.95. The correlation results suggest that there is less intraobserver error for the ribs than the correlation values for the pubic symphysis in Part I of this research.

Linear regression analyses were performed to compare the r values for the rib and the pubic symphysis in this study sample. The pubic symphysis r values for the males ($r = 0.867$) and females ($r = 0.877$) were less than the r values generated for the male rib ends ($r = 0.935$) and female rib ends ($r = 0.961$). The higher r values for the ribs suggest that the predictive values for the rib regression equations are much greater than those for the pubic symphysis, probably owing to the smaller age ranges for the ribs. The ribs showed less morphological variation than the pubic symphysis in each age group, contributing to the smaller age ranges.

The FSC phase descriptions and age ranges based on the sample collected at the Maricopa County Forensic Science Center in Phoenix, AZ, are modifications of those published by İşcan et al. (1–3). Minor changes were made to the wording, age ranges, and means per phase. The most important departure from the original study is the incorporation of bone quality and density as major factors in assigning age phases. In all, the morphological sort and seriation of the ribs in the FSC collection indicated that the variation among individuals of the same age and same age phase was much greater than that anticipated. The original proposal for this research aimed to reduce the size of the age ranges associated with the age phases. After the sort, the age ranges associated with specific morphological sets of features were, in fact, still fairly large. The original age ranges for stages of the sternal rib ends are wide for the older age groups, but they are wide for the purpose of taking into account this large amount of normal human variation.

Acknowledgments

Many thanks to the members of my committee, Dr. Brenda Baker (chair), Dr. Laura Fulginiti, Dr. Robert Williams, and Dr. Charles Merbs, for guiding me through this dissertation research. Many other individuals assisted me in this dissertation work, including all of the staff and doctors at the FSC. Additional specimens were obtained with help from the doctors and the staff of Barrow Neurological Institute and IIAM.

References

1. İşcan MY, Loth SR, Wright RK. Age estimation from the rib by phase analysis: white males. J Forensic Sci 1984;29:1094–104.
2. İşcan MY, Loth SR, Wright RK. Metamorphosis at the sternal rib end: a new method to estimate age at death in white males. Am J Phys Anthropol 1984;65:147–56.
3. İşcan MY, Loth SR, Wright RK. Age estimation from the rib by phase analysis: white females. J Forensic Sci 1985;30:853–63.
4. Bocquet-Appel JP, Masset C. Farewell to paleodemography. J Hum Evol 1982;12:321–33.
5. Konigsberg LW, Frankenberg SR, Walker RB. Regress what on what: paleodemographic age estimation as a calibration problem. In: Paine RR, editor. Integrating archaeological demography: multidisciplinary approaches to prehistoric population. Carbondale, IL: Southern Illinois University, 1994;64–88.
6. Jackes M. Building the bases for paleodemographic analysis: adult age determination. In: Katzenberg MA, Saunders SR, editors. Biological anthropology of the human skeleton. New York, NY: Wiley-Liss, Inc, 2000;417–66.
7. Boldsen JL, Milner GR, Konigsberg LW, Wood JW. Transition analysis: a new method for estimating age from skeletons. In: Hoppa RD, Vaupel JW, editors. Paleodemography: age distributions from skeletal samples. United Kingdom: Cambridge University Press, 2002;73–106.
8. Lucy D, Aykroyd RG, Pollard A, Solheim M. A Bayesian approach to adult human age estimation from dental observations by Johanson's age changes. J Forensic Sci 1996;41(2):189–94.
9. Konigsberg LW, Frankenberg SR. Deconstructing death in paleodemography. Am J Phys Anthropol 2002;117:297–309.
10. Aykroyd RG, Lucy D, Pollard AM, Solheim T. Technical note: regression analysis in adult age estimation. Am J Phys Anthropol 1997;104:259–65.
11. İşcan MY, Loth SR. Casts of age phases from the sternal end of the rib for white males and females. Fort Collins, CO: France Casting, 1993.
12. İşcan MY, Loth SR. Determination of age from the sternal rib in white females: a test of the phase method. J Forensic Sci 1986;31:990–9.
13. Loth SR. Age assessment of the Spitalfields cemetery population by rib phase analysis. Am J Hum Biol 1995;7:465–71.
14. Taylor KM. The effects of alcohol and drug abuse on the sternal end of the fourth rib [dissertation]. Tucson (AZ): University of Arizona, 2000.
15. Russell KF, Simpson SW, Genovese J, Kinkel MD, Meindl RS, Lovejoy CO. Independent test of the fourth rib aging technique. Am J Hum Biol 1993;92:53–62.
16. Fulginiti LC, Taylor K, Czuzak MH. Pubic symphysis takes a ribbing: comparison of age determination techniques in five forensic cases. Proceedings of the 47th Annual Meeting of the American Academy of Forensic Sciences; 1995 Feb 13-18; Seattle, WA. Colorado Springs, CO: American Academy of Forensic Sciences, 1995.
17. Fulginiti LC, Taylor K. Pubic symphysis takes a ribbing II. Proceedings of the 48th Annual Meeting of the American Academy of Forensic Sciences; 1996 Feb 19-24; Nashville, TN. Colorado Springs, CO: American Academy of Forensic Sciences, 1996.
18. Loth SR, İşcan MY. Skeletal aging techniques: the good, the bad, and the equivocal. Am J Phys Anthropol 1988;75:241.
19. Loth SR, İşcan MY. A systematic comparison of the accuracy of age estimation from the rib and pubis. Am J Phys Anthropol 1990;81:260.

Additional information and reprint requests:
Kristen M. Hartnett, Ph.D.
Office of Chief Medical Examiner
520 First Avenue
New York, 10016 NY
E-mail: khartnett@ocme.nyc.gov

Appendix A

Revised phase descriptions for sternal end of the fourth ribs for the FSC collection.

TABLE A1—*Revised fourth rib phase descriptions.*

Phase 1	The pit is shallow and flat, and there are billows in the pit. The pit is shallow U-shaped in cross-section. The bone is very firm and solid, smooth to the touch, dense, and of good quality. The walls of the rim are thick. The rim may show the beginnings of scalloping.
Phase 2	There is an indentation to the pit. The pit is V-shaped in cross-section, and the rim is well defined with round edges. The rim is regular with some scalloping. The bone is firm and solid, smooth to the touch, dense, and of good quality. There is no flare to the rim edges; they are parallel to each other. The pit is still smooth inside, with little to no porosity. In females, the central arc, which manifests on the anterior and posterior walls as a semicircular curve, is visible.
Phase 3	The pit is V-shaped, and there is a slight flare to the rim edges. The rim edges are becoming undulating and slightly irregular, and there may be remnants of scallops, but they look worn down. There are no bony projections from the rim. There is porosity inside the pit. The bone quality is good; it is firm, solid, and smooth to the touch. The rim edges are rounded, but sharp. In many females, there is a build-up of bony plaque, either in the bottom of the pit or lining the interior of the pit, creating the appearance of a two-layer rim. An irregular central arc may be apparent.
Phase 4	The pit is deep and U-shaped. The edges of the pit flare outwards, expanding the oval area inside the pit. The rim edges are not undulating or scalloped but are irregular. There are no long bony projections from the rim, and the rim edges are thin, but firm. The bone quality is good but does not feel dense or heavy. There is porosity inside the pit. In some males, two distinct depressions are visible in the pit. In females, the central arc may be present and irregular; however, the superior and inferior edges of the rim have developed, decreasing the prominence of the central arc.
Phase 5	There are frequently small bony projections along the rim edges, especially at the superior and inferior edges of the rim. The pit is deep and U-shaped. The rim edges are irregular, flared, sharp, and thin. There is porosity inside the pit. The bone quality is fair; the bone is coarse to the touch and feels lighter than it looks.
Phase 6	The bone quality is fair to poor, light in weight, and the surfaces of the bone feel coarse and brittle. There are bony projections along the rim edges, especially at the superior and inferior edges, some of which may be over 1 cm long. The pit is deep and U-shaped. The rim is very irregular, thin, and fragile. There is porosity inside the pit. In some cases, there may be small bony extrusions inside the pit. In females, the central arc is not prominent.
Phase 7	The bone is very poor quality, and in many cases, translucent. The bone is very light, sometimes feeling like paper, and feels coarse and brittle to the touch. The pit is deep and U-shaped. There may be long bony growths inside the pit. The rim is very irregular with long bony projections. In some cases, much of the cartilage has ossified and window formation occurs. In some females, much of the cartilage in the interior of the pit has ossified into a bony projection extending more than 1 cm in length.
Variant	In some males, the cartilage has completely or almost completely ossified. The ossification tends to be a solid extension of bone, rather than a thin projection. All of the bone is of very good quality, including the ossification. It is dense, heavy, and smooth. In these instances, bone quality should be the determining factor. There are probably other factors, such as disease, trauma, or substance abuse that caused premature ossification of the cartilage. When the individual is truly very old, the bone quality will be very poor. Be aware of these instances where a rib end may appear very old because of ossification of the cartilage but is really actually a young individual, which can be ascertained by bone quality. In these cases, consult other age indicators in conjunction with the rib end.

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01428.x
Available online at: interscience.wiley.com

PAPER

PHYSICAL ANTHROPOLOGY

Ginesse A. Listi,[1] *Ph.D.*

The Impact of Racial Metric Variation in the Os Coxae on the Morphological Assessment of Sex*

ABSTRACT: This study examines whether sex determination based on morphological traits in the os coxae is impacted by racial quantitative variation. Nineteen traits were evaluated independently by two observers in 876 os coxae. Chi-square test was used on a random sample of 400 individuals to assess whether the distribution of correct sex assessment varied for white and black individuals based on each trait individually and all 19 traits collectively, as well as on inter-observer agreement in correct sex assessment. Results indicate that accuracy of sex assessment varied between white and black individuals in certain individual traits; however, accuracy was not impacted when all traits were considered together. Furthermore, traits that showed significant variation in correct sex assessment between races generally were not related to size, but instead were "discrete." Finally, analyses of inter-observer variation suggest that disparities in sex assessment for some traits may be related to differences in trait interpretation between observers rather than morphological dissimilarities between races.

KEYWORDS: forensic science, forensic anthropology, racial quantitative variation, race determination, sex assessment, pelvic skeletal morphology

Assessment of the biological profile, including age, sex, and race (or ancestry) determination, is a basic objective in forensic anthropology. Among skeletal elements, the os coxae is of significant value for the anthropologist in meeting this objective. While considerable research has been undertaken establishing its usefulness for age and sex determination (1–9), fewer studies have explored its value in the assessment of race. Those that do exist generally have focused on quantitative methods. From these studies, researchers have been able to distinguish between white and black Americans based on certain measurements of the os coxae, portions of the os coxae, or the pelvis as a whole, with overall accuracies ranging from 57% to 97%. Additionally, many of these studies found that the pelves of black Americans tended to be "smaller" or "narrower" than those of white Americans, though such differences were not always statistically significant (10–17).

Determination of sex from the morphology of the pelvis is based on the premise that, in many of the traits, pelves of females are larger (longer, wider, or broader) than pelves of males. Yet, few studies have assessed whether the quantitative differences between white and black Americans affect, or are apparent in, the morphological traits used by anthropologists to determine the sex of the individual. In one notable exception, Patriquin et al. (2003) (18) found significant differences between white and black South Africans in the accuracy of morphological traits for sex determination.

The purpose of this study is to examine whether or not sex determination based on morphological traits is impacted by quantitative variation found in white and black Americans. Nineteen morphological traits in the os coxae were evaluated, and statistical analyses were used to assess variation between white and black Americans in the accuracy of sex determination based on all 19 traits collectively, on the accuracy of sex determination for each individual trait, and on inter-observer variation in correct sex assessment.

Materials and Methods

Two observers with comparable levels of education and experience independently assessed sex in 876 left os coxae. Data were collected from a sample comprising modern Americans of known age, sex, and race from three collections: the William M. Bass Donated Collection housed at the University of Tennessee, the Robert J. Terry Anatomical Skeletal Collection housed at the National Museum of Natural History, and the Donated Forensic Collection housed at Louisiana State University. All individuals included in this study were adults, meaning that the epiphyses were completely or nearly completely fused. Individuals with extreme degenerative changes or fusion of multiple elements were not assessed.

Nineteen morphological traits consisting of a combination of traditional and newer techniques were evaluated. While detailed descriptions can be found in previous research (1–9), all traits used in this study are summarized in Table 1 and include Phenice's traits, those relating to the shape of the pubic bone and to the size, shape, symmetry, and proportion of the greater sciatic notch, traits relating to the auricular surface and preauricular area, and traits of the inferior pelvis. Traits were scored as either male ("M"), female ("F"), or intermediate ("I"). The number of Ms, Fs, and Is then were counted, and the most numerous determined the overall sex

[1]Department of Geography and Anthropology, Louisiana State University, Baton Rouge, LA 70803, USA.

*Paper presented at the 61st Annual Meeting of the American Academy of Forensic Sciences, February 16–21, 2009, in Denver, CO, USA.

Received 9 Mar. 2009; and in revised form 12 June 2009; accepted 31 July 2009.

TABLE 1—*Morphological traits used for sex determination**.

	Trait	Male Form (“M”)	Female Form (“F”)	Other (“I”)
1	Ventral arc	Absent	Present	Intermediate
2	Subpubic angle	Convex/straight	Concave	Intermediate
3	Ridge on medial ischiopubic ramus	Absent	Present	Intermediate
4	Shape of pubic bone	Rectangular/narrow	Trapezoidal/broad	Intermediate
5	Greater sciatic notch (GSN) size	Deep	Shallow	Intermediate
6	GSN shape	Narrow	Wide	Intermediate
7	Auricular surface height	Flat	Raised	Intermediate
8	Preauricular sulcus (PAS)	Absent/slight grooves	Present/well-defined	Intermediate
11	PAS – “Negative relief”	None (smooth)/very slight	Deep depression w/pits	Intermediate
12	PAS – “Grooves/pitting”	Depression w/open circumference	Pits w/closed circumference	Intermediate
13	PAS – “Positive relief”	PAS Tubercle	No PAS tubercle	Intermediate
14	GSN – “Proportions”	Posterior chord < anterior chord	Anterior chord >= posterior chord	Intermediate
15	GSN – “Notch contour”	Asymmetrical	Symmetrical	Intermediate
16	GSN – “AP line”†	AP crosses into GSN	AP does not cross into GSN	Intermediate
17	Composite arch	Single curve	Double curve	Intermediate
18	External eversion	Absent	Present	Intermediate
19	Phallic ridge	Present	Absent	Intermediate
20	Robusticity	Robust	Gracile	Intermediate
21	Ischiopubic index	Pubic length < ischial length	Pubic length > ischial length	Intermediate

*Traits 1 through 8 after Rogers and Saunders (19); traits 11 through 21 after Bruzek (20).

†“AP line” refers to an imaginary line drawn from point “A” (defined as the top of the piriform tubercle [if present] or the anterior part of the posterior/inferior iliac spine) to point “P” (located perpendicular to “A” at the greatest depth of the GSN) (20).

of the individual. If the number of Ms and Fs was equal, the individual was scored as Indeterminate. For statistical analyses, “Indeterminate” was considered incorrect.

A random sample of 400 individuals was selected for statistical analyses, composed of 100 of each race and sex category. Males and females were analyzed separately; the mean age for each group was 45.5 and 53 years, respectively. A chi-square test was used to determine whether significant variation existed in the distribution of correct sex assessment between white and black individuals based on all traits collectively as well as for each trait individually. A chi-square test also was used to evaluate inter-observer agreement in correct sex assessment based on all traits collectively and for each trait individually. For statistical analysis, differences were considered significant if $p < 0.05$.

Hypotheses and Expectations

Hypotheses and expectations were generated to assist with the interpretation of data gathered in this study.

The first hypothesis is that the techniques for sex assessment will work equally well regardless of the race of the individual. If this is so, no significant differences in the distribution of correct sex assessment between white and black individuals are expected when all traits are considered either individually or collectively.

The second hypothesis is that sex determination based on traits that potentially are related to the size of the pelvis, or to the pelvic inlet or outlet, is more likely to vary between the races than sex estimation based on discrete traits. If so, significant differences in correct sex assessment are expected in traits that may be related to size (i.e., that are assessed descriptively as “narrow” or “wide/broad”), such as the shape of the pubic bone, greater sciatic notch, ischiopubic index, and perhaps, robusticity. Alternatively, traits that are assessed as “present/absent” would not show significant differences in correct sex assessment.

The third hypothesis is that, even though overall sex determination may not be impacted by race, the possibility that some traits will be affected will result in the disproportionate misclassification of sex in certain subgroups over others (specifically, in black females and white males). Therefore, the expectation is that sex in black females and white males will be misclassified more frequently than in white females and black males. This expectation is based on previous research on metric variation, which suggests that the pelves of black Americans tend to be “smaller” than the pelves of white Americans (10–17). As morphological sex determination is based on size (i.e., pelves of females are “larger” than pelves of males), the possibility of misclassification would be greater in the “small” pelves of black females and the “large” pelves of white males.

Results

Results indicate there was no significant difference between white and black Americans in the correct classification of sex for either observer when all traits were considered collectively. However, significant differences were noted in correct sex assessment for certain individual traits for both observers.

For Observer “A,” in males, 6/19 traits showed significant variation in the distribution of correct sex assessment between the races (Table 2). These traits included the medial aspect of the ischiopubic ramus, shape of the pubic bone, auricular surface height, composite arch, external eversion, and phallic ridge. In these, black males were classified correctly more often than white males in all traits except the composite arch. For females (Table 3), only one trait (composite arch) showed significant variation in the distribution of correct sex assessment, with black females being classified correctly more often than white females.

For Observer “B,” no significant differences in the distribution of correct sex assessment were found in any traits for males (Table 4). For females (Table 5), 3/19 traits showed significant variation in the distribution of correct sex assessment with white females being classified correctly more often than black females. These traits were all related to the preauricular sulcus and included its presence/absence and the existence of both grooves/pitting and negative relief.

Tables 6 and 7 present the results for the analyses of inter-observer agreement for males and females, respectively. Among males, 2/19 traits, including shape of the pubic bone and external eversion, showed significant variation in the agreement between

TABLE 2—*Correct sex assessment in males for Observer A.*

	Trait	Black	White	χ^2	*p* value
1	Ventral arc	88	91	0.479	0.489
2	Subpubic angle	96	90	2.765	0.096
3	Medial aspect/ischiopubic ramus	94	80	8.665	0.003*
4	Shape of pubic bone	96	86	6.105	0.013†
5	Greater sciatic notch (GSN) size	91	91	0.000	1.000
6	GSN shape	66	59	1.045	0.307
7	Auricular surface height	92	79	6.816	0.009*
8	Preauricular sulcus (PAS)	90	93	0.579	0.447
11	PAS – "Negative relief"	91	96	2.057	0.152
12	PAS – "Grooves/pitting"	96	98	0.687	0.407
13	PAS – "Positive relief"	84	78	1.170	0.279
14	GSN – Proportions	87	86	0.043	0.836
15	GSN – Notch contour	87	82	0.954	0.329
16	GSN – AP line	70	81	3.271	0.071
17	Composite arch	56	73	6.311	0.012†
18	External eversion	95	73	18.006	0.000‡
19	Phallic ridge	83	68	6.082	0.014†
20	Robusticity	87	86	0.043	0.836
21	Ischiopubic index	53	40	3.397	0.065
	Total	98	98	0.000	1.000

*Significant at $p < 0.01$; †Significant at $p < 0.05$; ‡Significant at $p < 0.001$.

TABLE 3—*Correct sex assessment in females for Observer A.*

	Trait	Black	White	χ^2	*p* value
1	Ventral arc	92	85	2.407	0.121
2	Subpubic angle	95	93	0.355	0.552
3	Medial aspect/ischiopubic ramus	91	88	0.479	0.489
4	Shape of pubic bone	95	93	0.355	0.552
5	Greater sciatic notch (GSN) size	72	78	0.960	0.327
6	GSN shape	94	88	2.198	0.138
7	Auricular surface height	89	87	0.189	0.663
8	Preauricular sulcus (PAS)	71	73	0.099	0.753
11	PAS – "Negative relief"	88	78	3.544	0.060
12	PAS – "Grooves/pitting"	70	69	0.024	0.878
13	PAS – "Positive relief"	78	72	0.960	0.327
14	GSN – Proportions	74	59	0.874	0.350
15	GSN – Notch contour	50	40	2.020	0.155
16	GSN – AP line	74	69	0.618	0.434
17	Composite arch	81	66	5.776	0.016*
18	External eversion	49	53	0.320	0.572
19	Phallic ridge	54	48	0.720	0.396
20	Robusticity	76	64	3.429	0.064
21	Ischiopubic index	95	96	0.116	0.733
	Total	95	94	0.096	0.756

*Significant at $p < 0.05$.

TABLE 4—*Correct sex assessment in males for Observer B.*

	Trait	Black	White	χ^2	*p* value
1	Ventral arc	69	80	3.185	0.074
2	Subpubic angle	78	76	0.113	0.737
3	Medial aspect/ischiopubic ramus	45	37	1.323	0.250
4	Shape of pubic bone	98	94	2.083	0.149
5	Greater sciatic notch (GSN) size	57	46	2.422	0.120
6	GSN shape	46	36	2.067	0.151
7	Auricular surface height	84	87	0.363	0.547
8	Preauricular sulcus (PAS)	93	94	0.082	0.774
11	PAS – "Negative relief"	96	99	1.846	0.174
12	PAS – "Grooves/pitting"	96	99	1.846	0.174
13	PAS – "Positive relief"	33	43	2.122	0.145
14	GSN – Proportions	92	89	0.267	0.605
15	GSN – Notch contour	89	83	1.130	0.288
16	GSN – AP line	60	58	0.041	0.839
17	Composite arch	55	50	0.403	0.525
18	External eversion	75	70	0.627	0.428
19	Phallic ridge	70	61	1.792	0.181
20	Robusticity	71	79	1.707	0.191
21	Ischiopubic index	91	82	3.468	0.063
	Total	98	99	0.338	0.561

TABLE 5—*Correct sex assessment in females for Observer B.*

	Trait	Black	White	χ^2	*p* value
1	Ventral arc	91	83	2.829	0.093
2	Subpubic angle	98	93	2.909	0.088
3	Medial aspect/ischiopubic ramus	82	77	0.767	0.381
4	Shape of pubic bone	60	73	3.793	0.051
5	Greater sciatic notch (GSN) size	27	29	0.099	0.753
6	GSN shape	88	85	0.385	0.535
7	Auricular surface height	74	81	1.405	0.236
8	Preauricular sulcus (PAS)	57	73	5.626	0.018*
11	PAS – "Negative relief"	55	71	5.491	0.019*
12	PAS – "Grooves/pitting"	52	69	6.047	0.014*
13	PAS – "Positive relief"	59	56	0.184	0.668
14	GSN – Proportions	53	54	0.020	0.887
15	GSN – Notch contour	56	43	3.380	0.066
16	GSN – AP line	83	74	2.400	0.121
17	Composite arch	73	79	0.987	0.321
18	External eversion	88	83	1.008	0.315
19	Phallic ridge	44	44	0.000	1.000
20	Robusticity	61	58	0.187	0.666
21	Ischiopubic index	85	86	0.040	0.841
	Total	90	95	1.802	0.179

*Significant at $p < 0.05$.

observers in correct sex assessment. Among females, significant inter-observer variation in the agreement between observers in correct sex assessment was found when all traits were considered collectively, as well as in 4/19 individual traits (i.e., ventral arc, and presence/absence, existence of grooves/pitting, and negative relief of the preauricular sulcus).

Discussion

The purpose of this study was to evaluate whether or not quantitative variation in the pelves of white and black Americans impacts the morphological assessment of sex. The first hypothesis, that the techniques for sex assessment would work equally well regardless of the race of the individual, generally, was supported. While some individual traits for each observer did show significant variation between races in the distribution of correct sex assessment, no significant differences were found when all traits were considered collectively. Additionally, correct sex classification when all 19 traits were considered collectively averaged 96% for both observers and ranged from a low of 90% in black females to a high of 99% in white males.

The second hypothesis, that traits which may be related to the size of the pelvis are more likely to be impacted by race than discrete traits, was not supported. Between the two observers, of the nine traits that showed significant variation in correct sex assessment, eight were discrete (that is, assessed as "present/absent"). Only the shape of the pubic bone (assessed as "wide" or "narrow") showed significant variation in correct sex classification in males for Observer A. These results seem to suggest that the quantitative variation between white and black individuals is not observable in, or cannot be explicitly associated with, morphological traits.

The third hypothesis, that correct sex classification would be disproportionate among the four subgroups, was partially supported. The expectation was that black females and white males would be

TABLE 6—*Inter-observer agreement in correct sex assessment in males.*

	Trait	Black	White	χ^2	*p* value
1	Ventral arc	74	83	2.400	0.121
2	Subpubic angle	78	79	0.030	0.063
3	Medial aspect/ischiopubic ramus	47	48	0.020	0.887
4	Shape of pubic bone	96	86	6.105	0.013*
5	Greater sciatic notch (GSN) size	59	51	1.293	0.256
6	GSN shape	61	55	0.739	0.390
7	Auricular surface height	82	75	1.452	0.228
8	Preauricular sulcus (PAS)	88	91	0.479	0.489
11	PAS – "Negative relief"	90	95	1.802	0.179
12	PAS – "Grooves/pitting"	93	98	2.909	0.088
13	PAS – "Positive relief"	41	45	0.326	0.568
14	GSN – Proportions	93	91	0.272	0.602
15	GSN – Notch contour	90	84	1.592	0.207
16	GSN – AP line	74	67	1.178	0.278
17	Composite arch	63	61	0.085	0.771
18	External eversion	76	59	4.367	0.037*
19	Phallic ridge	71	63	1.447	0.229
20	Robusticity	73	81	1.807	0.179
21	Ischiopubic index	57	46	2.422	0.120
	Total	98	99	0.338	0.561

*Significant at $p < 0.05$.

TABLE 7—*Inter-observer agreement in correct sex assessment in females.*

	Trait	Black	White	χ^2	*p* value
1	Ventral arc	93	81	6.366	0.012*
2	Subpubic angle	94	92	0.307	0.579
3	Medial aspect/ischiopubic ramus	81	76	0.020	0.887
4	Shape of pubic bone	64	75	2.854	0.091
5	Greater sciatic notch (GSN) size	43	33	2.122	0.145
6	GSN shape	85	84	0.036	0.845
7	Auricular surface height	70	79	2.132	0.144
8	Preauricular sulcus (PAS)	62	75	3.916	0.048*
11	PAS – "Negative relief"	57	77	9.046	0.003†
12	PAS – "Grooves/pitting"	58	73	4.978	0.026*
13	PAS – "Positive relief"	60	61	0.021	0.885
14	GSN – Proportions	73	82	2.323	0.128
15	GSN – Notch contour	68	65	0.202	0.635
16	GSN – AP line	78	83	0.795	0.372
17	Composite arch	78	77	0.029	0.866
18	External eversion	49	61	2.909	0.088
19	Phallic ridge	60	89	0.328	0.567
20	Robusticity	69	62	1.084	0.298
21	Ischiopubic index	87	90	0.442	0.506
	Total	89	97	4.916	0.027*

*Significant at $p < 0.05$; †Significant at $p < 0.01$.

misclassified more often than white females and black males. In fact, white males *were* misclassified more often than black males in 12/19 (63%) traits for both observers, including five of the six traits that showed significant differences between races in correct sex assessment. Among females, the results were more varied. For Observer A, the percentage of correct sex assessment actually was *higher* in black females in 15/19 (79%) traits, including the one that showed significant variation between races. For Observer B, the percentage of correct sex assessment was higher in 9/19 (47%) traits for black females, but also in 9/19 (47%) traits for white females. However, in the three traits that showed significant variation in sex assessment between races, black females *were* misclassified more often than white females. Thus, while white males were disproportionately misclassified (which supports hypothesis three), the results for females were less straightforward.

Additional support for hypothesis three comes from the results comparing inter-observer agreement in correct sex assessment. For males, there was greater agreement between observers in correct sex assessment for black males in 11/19 (58%) traits. For females, there was greater agreement between observers in correct sex assessment for white females not only in 11/19 (58%) individual traits but also for overall sex assessment, when all traits were considered collectively. These results suggest there may be greater variation in trait expression among black females and white males compared to black males and white females.

Although differences found in this study in the distribution of correct sex assessment between white and black Americans could be because of racial morphological dissimilarities, a second explanation must be considered. While none of the traits overlap between observers, five show significant *inter-observer* variation. Disagreement between observers could be because of ambiguity in the techniques used for assessing sex or the misapplication of the techniques by the observers. Regarding the former, an inherent subjectivity exists in the methodology because the variables assessed are qualitative rather than quantitative. Although the publication of "best practices" through the efforts of the Scientific Working Group for Forensic Anthropology (21) may help to mitigate such technical or methodological sources of inter-observer error, differences nevertheless will continue to exist between individuals in how traits are interpreted (i.e., what is "broad" to one observer may be "intermediate" or "narrow" to another). Regarding misapplication of techniques, observer unfamiliarity or inexperience with new or less commonly used methods increases the possibility of misidentification. Such errors can be avoided by remaining current and becoming proficient with new research methods and, if possible, honing skills through the practice of internal peer-review. At any rate, when all results from this study are taken into account, data suggest, at least for some of the traits, that inter-observer differences in trait interpretation may have contributed to the disparity in correct sex assessment between races.

In conclusion, while some individual traits showed significant variation in correct sex assessment between white and black Americans, these differences did not affect correct sex assessment when all traits were considered collectively and, furthermore, may be due, at least in part, to inter-observer differences in trait interpretation rather than to racial variation in trait expression. Finally, the objective of this research was not to be able to determine race from the pelvis, nor to propose which traits or combination of traits provide the "best method" for predicting sex. The purpose was to explore whether or not quantitative differences in the pelves of white and black Americans affect the morphological assessment of sex. Results from this study suggest that they do not.

Acknowledgments

The author acknowledges the individuals at the National Museum of Natural History, the University of Tennessee, and Louisiana State University, who allowed permission to study the collections in their care.

References

1. Işcan MY, Derrick K. Determination of sex from the sacroiliac joint: a visual assessment technique. Florida Sci 1984;47:94–8.
2. Phenice TW. A newly developed visual method of sexing the os pubis. Am J Phys Anthropol 1969;30:297–302.
3. St. Hoyme LE. Sex differentiation in the inferior pelvis. Coll Antropol 1984;8:139–53.
4. St. Hoyme LE, Iscan MY. Determination of sex and race: accuracy and assumptions. In: Iscan MY, Kennedy KAR, editors. Reconstruction of life from the skeleton. New York, NY: Wiley-Liss, Inc, 1989;53–93.
5. Sutherland LD, Suchey JM. Use of the ventral arc in pubic sex determination. J Forensic Sci 1991;36:501–11.

6. Lovejoy CO, Meindl RS, Mensforth R, Barton TJ. Chronological metamorphosis of the auricular surface of the ilium: a new method of the determination of adult skeletal age at death. Am J Phys Anthropol 1985;68:15–28.
7. Meindl RS, Lovejoy CO, Mensforth RP, Walker RA. A revised method of age determination using the os pubis with a review and tests of accuracy of other current methods of pubic symphyseal aging. Am J Phys Anthropol 1985;68:29–45.
8. Suchey JM, Katz D. Skeletal age standards derived from an extensive multiracial sample of modern Americans. Am J Phys Anthropol 1986;69:269.
9. Todd TW. Age changes in the pubic bone: I, the male white pubis. Am J Phys Anthropol 1920;3:285–334.
10. DiBennardo R, Taylor JV. Multiple discriminant function analysis of sex and race in the postcranial skeleton. Am J Phys Anthropol 1983;61:305–14.
11. Işçan MY. Assessment of race from the pelvis. Am J Phys Anthropol 1983;62:205–8.
12. Letterman GS. The greater sciatic notch in American whites and Negroes. Am J Phys Anthropol 1941;28:99–116.
13. Patriquin ML, Loth SR, Steyn M. Metric assessment of race from the pelvis in South Africans. Forensic Sci Int 2002;127:104–13.
14. Schulter-Ellis FP, Hayek LC. Predicting race and sex with an acetabulum/pubis index. Coll Antropol 1984;8:155–62.
15. Strauss WL. The human ilium: sex and stock. Am J Phys Anthropol 1927;XI:1–28.
16. Synstelien JA. Differences in the os coxae between blacks and whites: a musculoskeletal approach to human variation [thesis]. Knoxville (TN): Univ. of Tennessee, 2001.
17. Taylor JV, DiBennardo R. Discriminant function analysis on the central portion of the innominate. Am J Phys Anthropol 1984;1:315–20.
18. Patriquin ML, Loth SR, Steyn M. Sexually dimorphic pelvic morphology in South African whites and blacks. Homo 2003;53:255–62.
19. Rogers T, Saunders S. Accuracy of sex determination using morphological traits of the human pelvis. J Forensic Sci 1994;39:1047–56.
20. Bruzek J. A method for visual determination of sex using the human hip bone. Am J Phys Anthropol 2002;117:157–68.
21. http://swganth.org/ (accessed on June 12, 2009).

Additional information and reprints requests:
Ginesse A. Listi, Ph.D.
LSU Geography and Anthropology
227 Howe-Russell Building
Baton Rouge, LA 70803
E-mail: glisti1@lsu.edu

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01447.x
Available online at: interscience.wiley.com

PAPER

PHYSICAL ANTHROPOLOGY

Lelia Watamaniuk,[1] *B.Sc. and Tracy Rogers,*[1] *Ph.D.*

Positive Personal Identification of Human Remains Based on Thoracic Vertebral Margin Morphology

ABSTRACT: Radiography has long been used by anthropologists to establish positive personal identification of human remains in forensic cases. These methods have been largely *ad hoc* and depend upon specific congenital or pathological bone markers. Court rulings, such as Daubert and Mohan have, however, pushed the discipline toward more statistically supportable methods of identification. This study describes the use of normal morphological variation of the thoracic vertebrae to identify human remains. Radiographs from healthy, male individuals, aged 18–55 were examined to identify normally varying features of vertebral morphology. The frequency of occurrence of these features was calculated, tested, and found to be stable in the given sample. The frequencies were compared to establish which sets of traits varied independently of one another. Finally, unknown radiographs were compared to known samples to test the applicability of this method in determining positive identification, with 21 of 24 (87.5%) unknown radiographs positively identified.

KEYWORDS: forensic science, forensic anthropology, positive identification, human remains, thoracic, vertebrae, morphology, natural variation

The current fourth Era of Forensic Anthropology (1) is characterized by a contemplative and reflective approach to the discipline that has led to theory building, recommendations for best practices, and explicitly stated ethical positions. One aspect of this process is a new emphasis on validation studies, standardization of protocols, and mathematical statements of certainty, which originated in the broader field of forensic science in the wake of legal rulings on the admissibility of expert testimony in court (2). Both individual authors (3–6) and professional organizations, e.g., Scientific Working Group for Anthropology (swganth.org), are evaluating the assumptions, techniques, and approaches of forensic anthropology to provide experts a scientifically sound and legally admissible (7,8) means of selecting methods for use in forensic cases. Positive identification (ID) is one area for which verification is particularly critical because victim identification has significant investigative, legal, and emotional consequences (9).

Radiography has long been used to establish the identity of human remains when visual ID is impossible (10) and DNA is unavailable. Fractures, morphological features, and patterns of lesions unique to an individual have all been used to identify decomposed or skeletonized remains (11). The vast majority of these approaches, although widely accepted, are essentially *ad hoc* (12,13), lacking standardization and validation, and thus subject to dismissal under Daubert or Mohan. *Ad hoc* means of analysis are also scientifically limited, generating little data to aid in establishing their applicability to future cases. Given the prevalence of medical imaging and natural variation of the human skeleton, it is possible to document the frequencies of normal skeletal variants to devise standardized radiographic methods of establishing ID with known levels of certainty. Research into frontal sinuses and cranial sutures has demonstrated the unique nature of these traits (3,14), but their use in forensic cases is limited by the fact that only a small proportion of all radiographs taken for diagnostic purposes are from the skull. Brodgon (10) reports that antemortem thoracic radiographs represent 40% of the total number of radiographs taken during clinical examination, compared to 5% for cranial radiographs. Thus, an individual is more likely to have a chest X-ray taken during their lifetime than a radiograph from any other region of the body. With the high prevalence of thoracic radiographs in the antemortem record, it is logical to examine this aspect of the skeleton to develop a positive identification protocol.

The survivability of skeletal elements in a forensic context is another factor to consider when formulating a standardized method of positive skeletal ID. In cases of submersion, the vertebral column is among the last structures to be disarticulated and removed from the scene (15). For remains scattered above the ground, the cortical bone of the long bones and vertebrae survive longest in contexts without animal scavenging (16), and in the presence of several species of animals, the vertebral column is among the last structures to be disarticulated from the remains (17–21). Even in cases of deliberate postmortem disarticulation and alteration, e.g., burning or the removal of head or hands to obscure identity, the vertebral column is not generally disturbed (16).

Vertebral radiographs have been successfully used for positive identifications (12,22), but the procedures have proven difficult to quantify and standardize, and several significant problems have arisen, including: difficulty in orienting structures in postmortem radiographs to match those in antemortem orientations (11,12); visual interference by soft tissues obscuring relevant structures on

[1]Department of Anthropology, University of Toronto at Mississauga, 3359 Mississauga Rd. N. 212, NB, Mississauga, ON L5L 1C6, Canada.

Received 4 Oct. 2008; and in revised form 28 June 2009; accepted 10 July 2009.

the radiograph (11,12); the need for specific and potentially *ad hoc* metrics and measurements of thoracic features (22); and unknown magnification factors of the antemortem radiographs, making comparison with the postmortem specimen impossible. Finally, antemortem processes such as osteophytosis and postmortem taphonomic alterations may modify or obliterate landmarks used in metric measurements and comparisons (3,4). Given the difficulties associated with reliable, precise methods of quantitatively analyzing radiographs, it would seem a nonmetric scoring system, based on qualitative traits such as the size, shape, and presence or absence of a particular feature would be more useful to the investigator (4). It is just such a protocol that is examined in this study.

Materials and Methods

All chest radiographs used in the study were from men, aged 18–55. Both antero-posterior (AP) and lateral view radiographs were chosen from individuals who showed no evidence of disease, trauma, or malformation of the vertebral column. Radiographs were made available by Toronto General Hospital (TGH), under strict privacy conditions that limit the information available for each individual; thus, ancestry data is not available. A pilot study of 32 TGH radiographs was conducted to determine the most suitable vertebral traits for analysis. The seven margins of each vertebra observable on AP and lateral radiographs (superior, inferior, medial, lateral on AP view; and superior, inferior, and anterior on lateral view) exhibited distinct morphological variation and were easily visible in the pilot sample of thoracic vertebrae. The posterior margin was not examined in lateral view radiographs, as visual interference from the ribs, spinous, and transverse processes, and soft tissue consistently obscured this margin.

The middle two-thirds of each margin was examined to avoid potential age-related changes in morphology from natural processes such as arthritis and osteophytosis. The class of morphology was defined by the shape of the margin at the midline of the vertebra. In both the AP and lateral views, the superior and inferior margins showed five classes of regularly occurring morphology: flat (F), concave (C), wavy (W), discontinuous (D), and angled (A). Margins that did not demonstrate one of these five classes were scored as irregular (I). Figures 1–5 summarize these variations in morphology. Lateral margins (right, left, and anterior) in both the AP and lateral radiographic views showed six classes of regularly occurring morphology: straight vertical (SV), straight angular (SA), rough vertical (RV), rough angular (RA), concave (C), and pinched (P). Those margins that did not demonstrate one of these six morphologies were classified as irregular (I). Figures 6–11 illustrate these variants.

Having established the trait classes, location, and views for scoring in the pilot study, the frequency of each variant was established using an independent set of radiographs (n = 100), referred to as Group 1. For Group 1, each radiograph was examined and each vertebra was individually scored using the data sheet in Fig. 12. Radiographs in the AP view were oriented with the heart on the right as it faces the investigator. Scoring proceeded from the superior margin of T1 through T12; followed by the inferior margin of T1 through T12, the left lateral (heart side) margin from T1 through T12, and the right lateral (nonheart side) T1 to T12. The individual's lateral view was then examined and scored in the same fashion. The lateral view radiographs were oriented so that the anterior was toward the right and the posterior was toward the left when viewed by the investigator. Margins that were not clearly visible were not scored.

This process was repeated for each individual in Group 2 (a second independent sample, n = 100) to confirm the frequencies observed in Group 1. In order to avoid potential bias for Group 2 results, the frequencies of Group 1 variants were not calculated

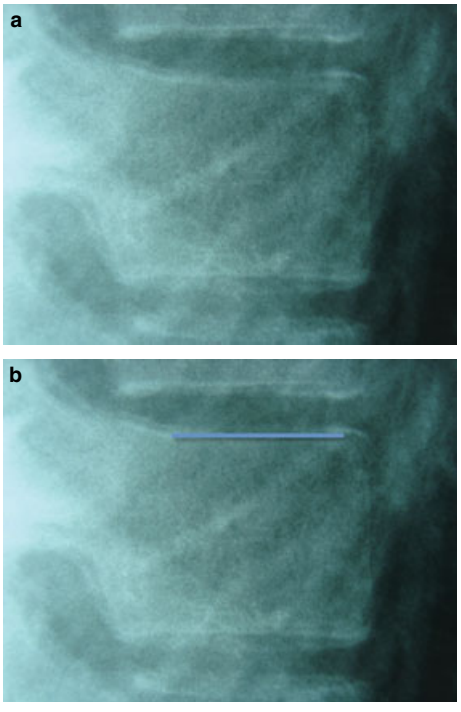


FIG. 1—*(a) Flat (F): Original photo. The margin remains flat and horizontal across the sagittal midline. (b) Flat (F): Modified photo emphasizing the flat (F) margin.*

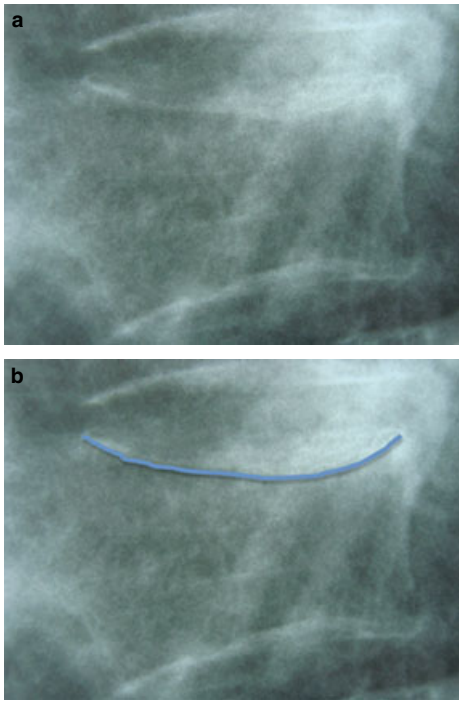


FIG. 2—*(a) Concave (C): The margin curves in toward the center of the vertebra at the sagittal midline. (b) Concave (C): Modified photo emphasizing the concave (C) margin.*

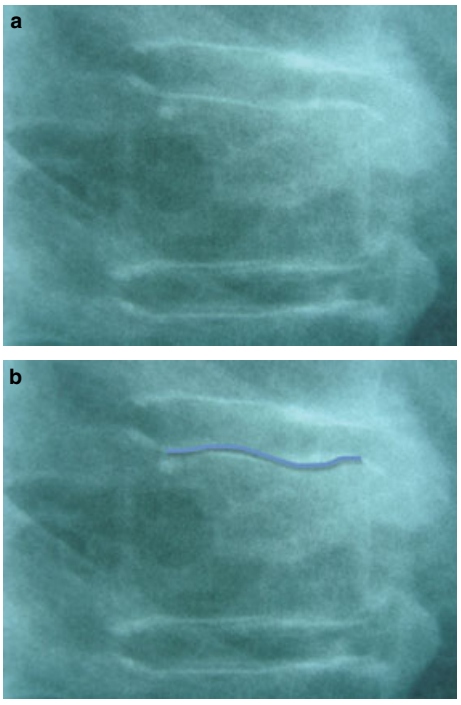


FIG. 3—*(a) Wavy (W): The margin curves toward and then away from the center of the vertebra at the sagittal midline. (b) Wavy (W): Modified photo emphasizing the wavy (W) margin.*

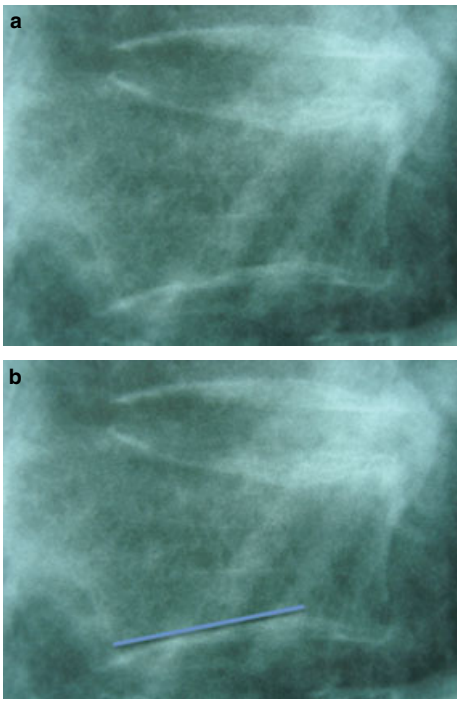


FIG. 5—*(a) Discontinuous (D): The margin is interrupted across the medial two-thirds of the vertebra. (b) Discontinuous (D): Modified photo emphasizing the discontinuous (D) margin.*

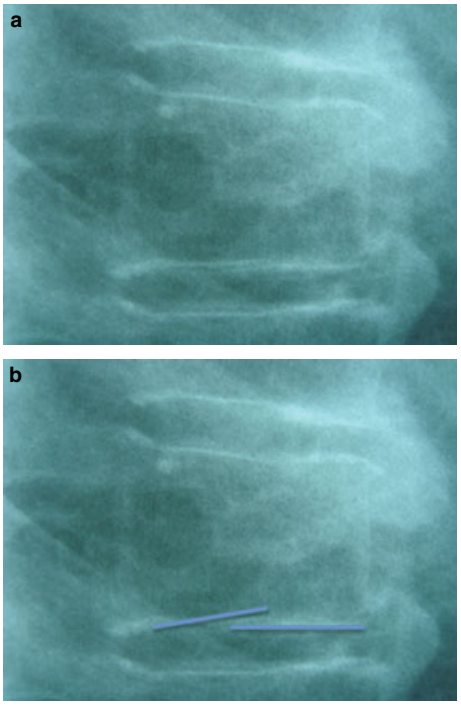


FIG. 4—*(a) Angled (A): The margin forms an angle across the sagittal midline. (b) Angled (A): Modified photo emphasizing the angled (A) margin.*

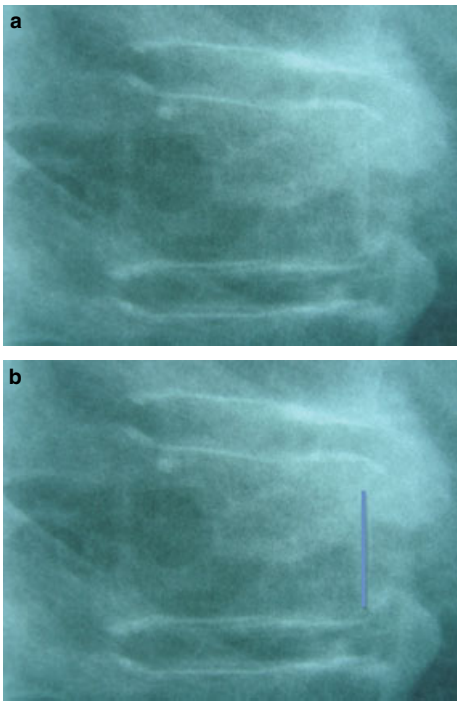


FIG. 6—*(a) Smooth Vertical (SV): The margin remains smooth and is not angled across the transverse midline. (b) SV: Modified photo emphasizing the SV margin.*

until after Group 2 radiographs were scored. Once Group 2 scoring was complete, variant frequencies for both Group 1 and Group 2 were calculated and compared using chi-square analysis. Trait independence was evaluated to determine if the shape of a vertebral margin was dependent upon other margins in the same vertebra, and/or its position within the vertebral column, i.e., all superior margins had the same shape, etc. Scoring was repeated on Group 1 radiographs at a later date by the same investigator to calculate intra-observer error and by a second observer to calculate inter-observer error.

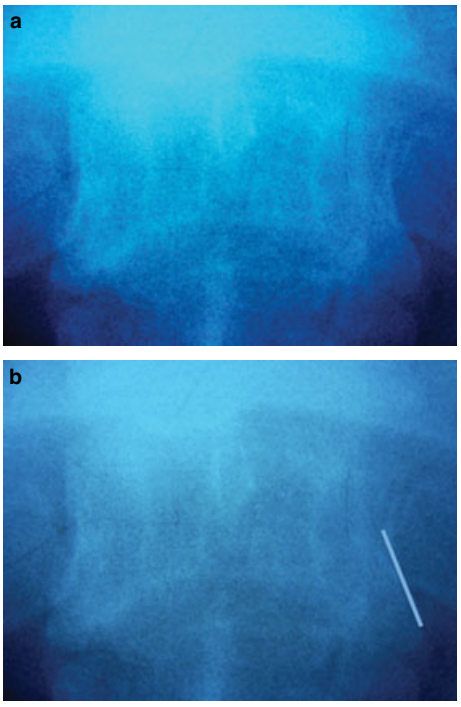


FIG. 7—*(a) Rough Vertical (RV): The margin is not smooth in appearance and is not angled across the transverse midline. (b) RV: Modified photo emphasizing the RV margin.*

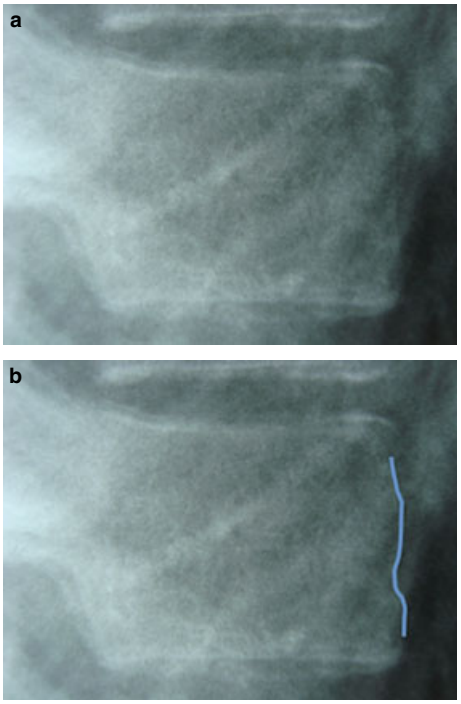


FIG. 8—*(a) Smooth Angled (SA): The margin remains smooth and is angled across the transverse midline. (b) SA: Modified photo emphasizing the SA margin.*

Finally, Group 3 (a third independent set of radiographs, $n = 100$) was used to determine whether the calculated variant frequencies could be used to establish the identity of a set of "unknown" radiographs. Duplicate radiographs for 24 of the 100 individuals in Group 3 were selected and designated "unknowns." The identity of the "unknowns" was obscured and hidden from the examiner. Each of the 24 unknowns was compared to the original sample of 100 in an attempt to match it to a known individual using vertebral morphology. Based on the frequencies established using Group 1, the strength of the identification was then calculated and the identification was classed as either positive or tentative.

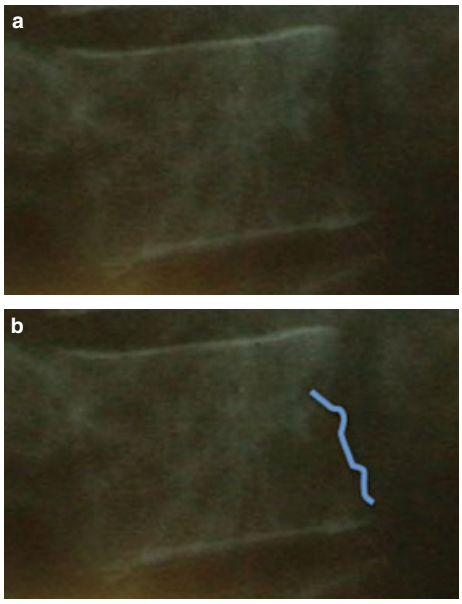


FIG. 9—*(a) Rough Angled (RA): The margin is not smooth in appearance and is angled across the transverse midline. (b) RA: Modified photo emphasizing the RA margin.*

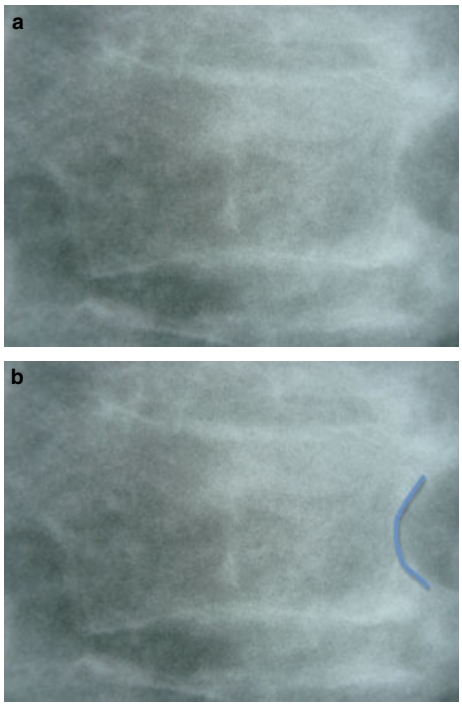


FIG. 10—*(a) Concave (C): The margin curves in toward the center of the vertebra at the transverse midline. (b) Concave (C): Modified photo emphasizing the concave (C) margin.*

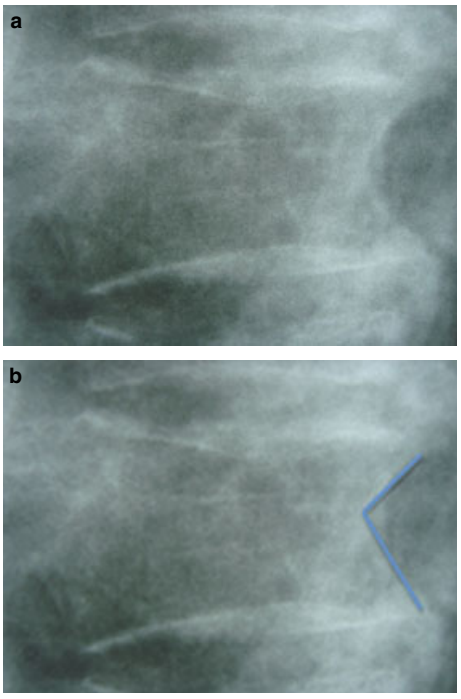


FIG. 11—*(a) Pinched (P): The margin forms a "V" with the apex pointing toward the center of the vertebra at the transverse midline. (b) Pinched (P): Modified photo emphasizing the pinched (P) margin.*

Results

The frequencies of each variant of each margin of each vertebra were calculated. Scoring vertebra above the sixth thoracic in the AP view was possible in fewer than 15% of the radiographs examined. Soft tissue of the chest, such as the lungs and heart, obscured the margins of the vertebrae, in particular, the left lateral margin, making morphological assessment difficult. For this reason, frequencies for the AP views for T1–T5 are not included in Tables 1 and 2. Similarly, the margins of T1–T5 are obscured in roughly 70% of the lateral radiographs of Group 1, in this case by ribs, and are therefore omitted from the summaries. Although each of the seven margins of T6 and T7 show some degree of visual interference, the morphology of these margins is far more easily discerned, and scores from both vertebrae were included in the final results. The soft tissue of the diaphragm occasionally obscured the margins of T12, although not to the same degree as was observed with the upper thoracic vertebrae. In general, visibility and "scorability" were found to be more straightforward on the lateral view radiographs.

Both intra- and inter-observer error were calculated for each margin of each vertebra from T6 to T12. Intra-observer error is the percentage of errors made between the original scores and the subsequent scores made by the researcher, while inter-observer error is the percentage of errors made between the first and second researcher for each vertebral margin. Intra-observer error was found to be low, with an average value of 3.4% across the vertebral margin sites. The range of intra-observer error was 4–20%. The highest instances of intra-observer error, 20% and 18%, occurred on the right margin of T6 and the left margin of T7, respectively, in the AP view. Both instances of high error were observed on vertebral margins where morphology was not readily scored because of significant visual interference from soft tissue. Inter-observer error was problematic (average error 33% across vertebral margins). Again the highest rates of error were found on margins obscured by soft tissue, for instance, the right lateral margins viewed on the AP aspect radiographs showed the highest rate of error (average 46%; range 17–100%), as well as most margins on the T6 and T7 vertebrae. The high rates of inter-observer error are believed to be because of researcher instruction and will be addressed in subsequent studies.

Frequencies for each class of morphological variation (5 or 6 depending on view) were calculated for each of the seven margins on each vertebra (T6–T12). Group 2 frequencies were compared to corresponding frequencies of Group 1 to determine if substantial variation existed. Chi-square analysis (d.f. = 1, $p \leq 0.05$, $\chi^2 = 3.84$) confirmed that with the exception of those frequencies discussed later, no significant differences existed between those of Group 1 and Group 2. Seven frequencies showed significant differences using chi-square analysis. Two occurred along the left lateral margins and two on the right in the AP view, which typically show soft tissue interference; the remaining three were along superior margins in the lateral view on the T10 and T11 vertebrae. The instances of significant differences in the latter three may have been caused by soft tissue interference, or given that they were in the lateral view, it is possible that ribs may have obscured or interfered with the scoring in these regions.

Complex patterns of trait independence were found when the variant frequencies were examined between margins of a single vertebra, and when the same margin was compared across vertebrae (i.e., all superior margins from T6 to T12). Inter-vertebrally, most margins demonstrated a dependent pattern of variant occurrence; meaning, for example, that the variant observed on the superior margin of T6 was most likely the same one found on the superior margins of T7–T12. Only five variants showed independent patterning across the vertebral column: SV along the left margin; RV and C along the anterior lateral margin; C and I along the superior margin in the AP view. Intra-vertebrally, certain variants display independence across margins. The pattern is quite complex, but the trend appears to be that the lower vertebrae (T10–T12) demonstrate greater independent variation than the mid thoracic vertebrae (T6–T9). This suggests that, perhaps because of the greater load-bearing function of the lower vertebrae, noncongenital, nondevelopmental remodeling may be a source of variation.

Testing this technique for use in positive identification was undertaken on the basis of visual inspection and comparison of 24 radiographs of "unknown" (as described previously) individuals to a sample of 100 "known" sets of radiographs. To determine the probability of achieving a unique or positive ID, the frequencies of independent variants were multiplied together. For the known sample of 100, if the product of frequencies was <0.01 (<1 in 100), the identification was considered positive, if it was >0.01, the identification was considered "possible," and those probabilities calculated to be near 0.01 would be considered tentative. For example, with a probability calculated to be 0.0370, radiograph #10 could only be identified as a "possible" match, as it is possible that the individual shared a similar pattern of vertebral traits with four individuals in the comparative group of 100.

Of the 24 unknown radiographs compared against the known sample, 21 were positively identified (Table 3). Of the three that were not identified positively, two of the unknown radiographs, #2 and #19 were of a poor quality, i.e., were over exposed with poor resolution of the vertebral margins, such that meaningful comparison with the known sample was not possible. No attempt at identification was possible—an incorrect ID was not made, but neither was a correct ID possible. The third radiograph, #10, could not be positively identified. In this case, the majority of vertebral margins

VERTEBRAL MORPHOLOGY FOR POSITIVE IDENTIFICATION

DATE: _____ RESEARCHER: _____

INDIVIDUAL NUMBER: _____ INDIVIDUAL SEX: M/F AGE: _____

ANTERIOR/POSTERIOR VIEWS RADIOGRAPH NUMBER:

VERTEBRAL MARGINS:

	T6	T7	T8	T9	T10	T11	T12
Superior Margin							
Inferior Margin							

F -flat; C - concave; X - convex; W - wavy; D - discontinuous; A - angled; "/" - absent/unreadable

	T6	T7	T8	T9	T10	T11	T12
Left Lateral Margin							
Right Lateral Margin							

SV - smooth vertical; SA - smooth angular; RV - rough vertical; RA - rough angular; C - concave; P - pinched; I - irregular; "/" - absent/unreadable

LATERAL VIEW RADIOGRAPH NUMBER:

VERTEBRAL MARGINS:

	T6	T7	T8	T9	T10	T11	T12
Superior Margin							
Inferior Margin							

F -flat; C - concave; X - convex; W - wavy; D - discontinuous; A - angled; "/" - absent/unreadable

	T6	T7	T8	T9	T10	T11	T12
Anterior Margin							

SV - smooth vertical; SA - smooth angular; RV - rough vertical; RA - rough angular; C - concave; P - pinched; I - irregular; "/" - absent/unreadable

NOTES:

Watamaniuk 2010

FIG. 12—*Data sheet for scoring thoracic vertebral margin morphology.*

TABLE 1—*Superior and inferior margin morphological variant frequencies.*

Antero-posterior (AP) view							
Superior margin frequencies	T6	T7	T8	T9	T10	T11	T12
Flat (F)	0.44	0.64	0.67	0.77	0.89	0.93	0.93
Concave (C)	0.56	0.36	0.30	0.21	0.11	0.08	0.03
Wavy (W)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Discontinuous (D)	0.00	0.00	0.03	0.01	0.00	0.00	0.00
Angled (A)	0.00	0.00	0.00	0.01	0.00	0.00	0.00
Irregular (I)	0.00	0.00	0.00	0.00	0.00	0.00	0.03
Inferior margin frequencies	T6	T7	T8	T9	T10	T11	T12
Flat (F)	0.65	0.60	0.63	0.62	0.63	0.58	0.43
Concave (C)	0.35	0.40	0.34	0.35	0.35	0.41	0.53
Wavy (W)	0.00	0.00	0.00	0.04	0.00	0.00	0.00
Discontinuous (D)	0.00	0.00	0.03	0.00	0.00	0.00	0.03
Angled (A)	0.00	0.00	0.00	0.00	0.00	0.02	0.03
Irregular (I)	0.00	0.00	0.00	0.00	0.02	0.00	0.00
Lateral view							
Superior margin frequencies	T6	T7	T8	T9	T10	T11	T12
Flat (F)	0.71	0.77	0.66	0.66	0.62	0.73	0.77
Concave (C)	0.24	0.14	0.24	0.25	0.23	0.24	0.19
Wavy (W)	0.02	0.05	0.03	0.04	0.06	0.00	0.00
Discontinuous (D)	0.02	0.03	0.04	0.02	0.02	0.01	0.02
Angled (A)	0.02	0.02	0.03	0.04	0.05	0.01	0.03
Irregular (I)	0.00	0.00	0.00	0.00	0.01	0.00	0.00
Inferior margin frequencies	T6	T7	T8	T9	T10	T11	T12
Flat (F)	0.67	0.60	0.58	0.49	0.38	0.43	0.56
Concave (C)	0.23	0.21	0.26	0.40	0.44	0.35	0.28
Wavy (W)	0.09	0.10	0.05	0.03	0.14	0.12	0.10
Discontinuous (D)	0.01	0.06	0.08	0.07	0.02	0.05	0.02
Angled (A)	0.00	0.04	0.02	0.01	0.01	0.04	0.05
Irregular (I)	0.00	0.00	0.00	0.00	0.00	0.01	0.00

TABLE 2—*Lateral margin morphological variant frequencies.*

Antero-posterior (AP) view							
Left lateral margin frequencies	T6	T7	T8	T9	T10	T11	T12
Straight vertical (SV)	1.00	0.27	0.08	0.08	0.02	0.06	0.13
Rough vertical (RV)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Straight angular (SA)	0.00	0.18	0.20	0.26	0.46	0.38	0.15
Rough angular (RA)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Concave (C)	0.00	0.46	0.52	0.56	0.39	0.51	0.53
Pinched (P)	0.00	0.09	0.20	0.10	0.14	0.04	0.19
Irregular (I)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Right lateral margin frequencies	T6	T7	T8	T9	T10	T11	T12
Straight vertical (SV)	0.14	0.10	0.15	0.12	0.15	0.09	0.05
Rough vertical (RV)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Straight angular (SA)	0.14	0.13	0.19	0.19	0.34	0.33	0.36
Rough angular (RA)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Concave (C)	0.14	0.65	0.49	0.51	0.38	0.40	0.41
Pinched (P)	0.57	0.13	0.17	0.18	0.14	0.18	0.18
Irregular (I)	0.00	0.00	0.00	0.01	0.00	0.00	0.00
Lateral view							
Anterior lateral margin frequencies	T6	T7	T8	T9	T10	T11	T12
Straight vertical (SV)	0.54	0.66	0.48	0.54	0.44	0.37	0.33
Rough vertical (RV)	0.20	0.04	0.19	0.21	0.21	0.07	0.04
Straight angular (SA)	0.03	0.06	0.01	0.04	0.09	0.09	0.08
Rough angular (RA)	0.03	0.00	0.00	0.00	0.02	0.00	0.02
Concave (C)	0.17	0.23	0.27	0.20	0.24	0.46	0.50
Pinched (P)	0.03	0.00	0.05	0.00	0.02	0.01	0.02
Irregular (I)	0.00	0.00	0.00	0.00	0.00	0.00	0.00

demonstrated the high-frequency variants, e.g., flat superior and inferior borders and concave lateral borders, which meant that the "unknown" matched a number of potential "knowns." The correctly identified individuals were identified primarily on the basis of variants that occur in low frequencies, making the combination of traits unique to the individual in the comparative group.

Discussion and Conclusion

It is not always possible to rely upon unique skeletal markers when attempting to establish positive personal identification of remains. The individual in question may not have suffered a lesion, trauma, or other specific congenital condition that would uniquely mark their bones, nor would their antemortem radiographs necessarily record such an event. The use of cranial features such as frontal sinus patterns may be precluded by a lack of antemortem radiographs. The widespread availability of antemortem chest radiographs and the high likelihood of recovering thoracic vertebrae in forensic contexts make this region of the skeleton an important focus for research and techniques of positive identification. The current research confirmed: (i) the natural variability of the vertebral margins; (ii) the fact that the variants occur at consistent frequencies between two sample groups of men aged 18–55, suggesting these values are stable frequencies; (iii) that it is possible to devise a mathematically sound method of determining the likelihood of establishing a positive ID based on this variation; and (iv) that the proposed technique was 100% successful when it could be applied to this sample.

The other advantage of this technique is that the frequency data allow researchers to calculate the strength of a "potential" match. Determining which variants of vertebral margin morphology occur independent of one another is crucial to this process. Independent variant frequencies may be multiplied together to establish the probability of that specific combination of traits occurring together, e.g., three in 100 people exhibit that combination. Dependent variants, those whose occurrence predicts the occurrence of the same variant at a different location, must be treated as a single entity for all calculations. Once the probability is calculated, the match is made on the basis of achieving a probability that exceeds the sample size, in this case (where the sample size is 100), the trait frequencies must have a probability of 1/100 or less to be considered a positive match. To utilize this method in actual forensic cases, the calculated probabilities must be lower than 1 in whatever the size of the "Identification Universe" (the group to which the unknown individual will be compared in order to make an identification, e.g., the missing persons list, or those missing persons whose demographic information match the skeletal biological profile).

In a forensic context, unknown human remains are typically compared to a list of missing persons. For example, in 2002, there were more than 4000 missing persons in Ontario, Canada. An analysis of the biological profile of found human remains provides information about the sex, age, ancestry, and stature that allows the police, coroner/medical examiner, and forensic anthropologist to eliminate a large proportion of the missing persons list. In the case of Ontario, a list of more than 4000 can potentially be reduced to a few hundred by matching the demographic information of the missing person to the skeletal biological profile. Thus, the relevant comparative sample, or Identification Universe, would be those few hundred that are not eliminated by the biological profile. Depending on the variants observed on the thoracic vertebrae, it might be possible to establish a positive ID at this stage. Rare combinations will occur in frequencies lower than the size of the Identification Universe. For more common trait combinations, a possible ID with a statement of the probability can be achieved. For example, if the combination of traits found in the remains occurs at a frequency of

TABLE 3—*Summary of unknown identifications and positive identification probabilities.*

Unknown	Individual Matched To	Actual Identity	Margins Scored	Variant	Frequency	Probability
1	185	185	T10 AP left	SA	0.46	0.0048
			T10 lateral anterior	RV	0.21	
			T10 lateral superior	A	0.05	
2	No match	252	**No score, poor radiograph			
3	299	299	T8 lateral anterior	RV	0.19	0.0007
			T9 AP right	P	0.18	
			T12 lateral anterior	P	0.02	
4	137	137	T9 lateral inferior	A	0.01	0.0035
			T9 AP inferior	C	0.35	
5	189	189	T10 AP left	C	0.39	0.0023
			T12 AP left	SA	0.53	
			T12 AP superior	F	0.93	
			T11 AP inferior	F	0.58	
			T7 lateral superior	C	0.14	
			T8 lateral anterior	C	0.27	
			T9 lateral anterior	SV	0.55	
6	267	267	T11 lateral superior	A	0.01	0.0010
			T7 lateral inferior	W	0.10	
7	247	247	T8 lateral anterior	SV	0.48	0.0053
			T9 lateral anterior	C	0.20	
			T10 lateral anterior	C	0.24	
			T11 lateral anterior	C	0.46	
			T12 lateral anterior	C	0.50	
8	294	294	T6 lateral anterior	C	0.17	0.0058
			T7 lateral anterior	C	0.23	
			T8 lateral anterior	C	0.27	
			T9 lateral anterior	SV	0.55	
9	249	249	T8 lateral anterior	SV	0.48	0.0071
			T10 lateral anterior	RV	0.21	
			T11 lateral anterior	RV	0.07	
10	287	075	T9 AP superior	F	0.77	0.0370
			T9 lateral anterior	C	0.20	
			T10 lateral anterior	C	0.24	
11	231	231	T10 lateral inferior	A	0.01	0.0010
			T12 lateral inferior	W	0.10	
12	239	239	T10 lateral inferior	F	0.38	0.0084
			T10 lateral anterior	SV	0.44	
			T11 lateral inferior	D	0.05	
13	284	284	T9 lateral inferior	W	0.02	0.0036
			T9 AP right	P	0.18	
14	354	354	T10 lateral inferior	W	0.14	0.0054
			T10 lateral anterior	RV	0.21	
			T12 lateral anterior	C	0.50	
			T11 lateral anterior	SV	0.37	
15	019	019	T9 lateral inferior	A	0.01	0.0021
			T10 lateral anterior	RV	0.21	
16	154	154	T9 AP right	SA	0.19	0.0019
			T11 lateral superior	A	0.01	
17	327	327	T8 AP left	P	0.20	0.0005
			T7 AP right	P	0.13	
			T10 lateral inferior	D	0.02	
18	163	163	T10 lateral inferior	W	0.14	0.0013
			T8 lateral anterior	SV	0.48	
			T9 lateral anterior	RV	0.21	
			T10 lateral anterior	SA	0.09	
19	No match	324	**No score, poor radiograph			
20	318	318	T7 lateral superior	D	0.03	0.0072
			T8 lateral superior	C	0.24	
21	319	319	T8 later inferior	W	0.05	0.0001
			T9 lateral inferior	W	0.02	
			T10 lateral inferior	W	0.14	
22	321	321	T9 AP left	SA	0.26	0.0021
			T11 AP left	SA	0.38	
			T12 AP left	SA	0.15	
			T10 AP right	P	0.14	
23	287	287	T8 AP left	SA	0.20	0.0011
			T10 AP right	SA	0.34	
			T11 AP left	SA	0.38	
			T12 lateral anterior	RV	0.04	
24	217	217	T10 lateral anterior	SV	0.44	0.0004
			T6 lateral inferior	C	0.23	
			T10 lateral superior	W	0.06	
			T11 lateral anterior	RV	0.07	
					Positive ID > 0.010	

RV, rough vertical; SA, straight angular; SV, straight vertical

0.003, and the biological profile limits the potential matches to 345 individuals, the probability of the traits occurring together (3/1000) means only 1.035 people would be expected to have that combination in a group of 345 individuals. Lack of a match in a given Identification Universe or the expansion of an investigation may lead to comparisons with another group of missing persons. In this case, the identification protocol is the same, but the calculated probability of a match would be compared to the size of the new Identification Universe.

Worth noting is the potential for this protocol to eliminate the need for maceration in cases where it is impractical to clear soft tissue from bones. In cases of immolation, removal of soft tissue may be complicated by organic and chemical contamination of the remains, which may expose examiners to biohazardous material. Such cases would require additional resources and time, which may not be available to investigators. A protocol that reduces an investigator's dependence on maceration would save both time and resources in the identification of human remains.

The pattern of independent and dependent variants was not easily predictable, which suggests that frequency tables should be established and verified for a variety of populations to either confirm the stability of the frequencies presented here, or demonstrate the need for population specific frequencies. Ancestry was unknown in this study, but the nature of the radiographic collection, stemming from a large multicultural urban center, suggests the sample was biologically and ethnically diverse. Being a preliminary study, additional tests of the frequency data and this approach to positive ID will be necessary for women, children, the elderly, groups of known ancestry, and groups of individuals engaged in different occupations, e.g., "blue" versus "white" collar workers. The results of such future analyses will be invaluable for the study of human identification, but may also yield interesting information regarding sex, age, and ancestry, leading to new techniques for establishing a biological profile.

References

1. Sledzik PS, Fenton TW, Warren MW, Byrd JE, Crowder C, Drawdy SM, et al. The fourth era of forensic anthropology. Proceedings of the 59th Annual Meeting of the American Academy of Forensic Sciences; 2007 Feb 19–24; San Antonio, TX. Colorado Springs, CO: American Academy of Forensic Sciences, 2007;13:350–3.
2. United States of America v. Plaza, Acosta and Rodriguez, Cr. No. 98-362-10, 11, 12. United States District Court for the Eastern District of Pennsylvania, January 7, 2002.
3. Christensen AM. The impact of *Daubert*: implications for testimony and research in forensic anthropology (and the use of frontal sinuses in personal identification). J Forensic Sci 2004;49(3):427–30.
4. Rogers TL, Allard TT. Expert testimony and positive identification of human remains through cranial suture patterns. J Forensic Sci 2004; 49(2):203–7.
5. Grivas CR, Komar DA. Kumho, *Daubert*, and the nature of scientific inquiry: implications for forensic anthropology. J Forensic Sci 2008; 53(4):771–6.
6. Rogers TL. Determining the sex of human remains through cranial morphology. J Forensic Sci 2005;50(3):493–500.
7. *Daubert v. Merrell Dow Pharmaceuticals* 1993;113 S. Ct. 2786.
8. *Regina v. Mohan* [1994] 2.S.C.R. 19994;File No.:23036.
9. Haglund W. The National Crime Information Center (NCIC) missing and unidentified persons system revisited. J Forensic Sci 1993;38:365–78.
10. Brodgon BG. Forensic radiology. Boca Raton, FL: CRC Press, 1998.
11. Murphy WA, Spruill FG, Gantner GE. Radiologic identification of unknown human remains. J Forensic Sci 1980;25(4):727–35.
12. Kahana T, Goldin L, Hiss J. Personal identification based on radiographic vertebral features. Am J Forensic Med Pathol 2002;23(1):36–41.
13. Jablonski NG, Shum BS. Identification of unknown human remains by comparison of antemortem and postmortem radiographs. Forensic Sci Int 1989;42(3):221–30.
14. Sekharan PC. The individual characteristics of ectocranial sutures. Ind J Forensic Sci 1987;1:75–91.
15. Haglund WD. Disappearance of soft tissue and the disarticulation of human remains from aqueous environments. J Forensic Sci 1993;38(4):806–15.
16. Galloway A, Birkby WH, Jones AM, Henry TE, Parks BO. Decay rates of human remains in an arid environment. J Forensic Sci 1989;34(3):607–16.
17. Carson EA, Stefan VH, Powell JF. Skeletal manifestations of bear scavenging. J Forensic Sci 2000;45(3):515–26.
18. Haglund WD, Reay DT, Swindler DR. Tooth marks artifacts and survival of bones in animal scavenged human skeletons. J Forensic Sci 1988;33(4):985–97.
19. Haglund WD, Read DT, Swindler DR. Canid scavenging/disarticulation sequence of human remains in the Pacific Northwest. J Forensic Sci 1989;34(3):587–606.
20. Haglund WD. Contribution of rodents to postmortem artifacts of bone and soft tissue. J Forensic Sci 1992;37(6):1459–65.
21. Willey P, Snyder LM. Canid modification of human remains: implications for time since death estimations. J Forensic Sci 1989;34(4):894–901.
22. Kuehn CM, Taylor KM, Mann FA, Wilson AJ, Harruff RC. Validation of chest X-ray comparisons for unknown decedent identification. J Forensic Sci 2002;47(4):725–9.

Additional information and reprint requests:
Lelia Watamaniuk, B.Sc.
Department of Anthropology
University of Toronto at Mississauga
3359 Mississauga Rd. N. 212, NB
Mississauga, ON L5L 1C6
Canada
E-mail: l.watamaniuk@utoronto.ca

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01446.x
Available online at: interscience.wiley.com

PAPER

CRIMINALISTICS; GENERAL

James M. Curran,[1] *M.Sc.(Hons.), Ph.D. and John Buckleton,*[2] *Ph.D.*

Inclusion Probabilities and Dropout

ABSTRACT: Recent discussions on a forensic discussion group highlighted the prevalence of a practice in the application of inclusion probabilities when dropout is possible that is of significant concern. In such cases, there appears to be an unpublished practice of calculation of an inclusion probability only for those loci at which the profile of interest (hereafter the suspect) is fully included among the alleles present in the crime scene sample and to omit those loci at which the suspect has alleles that are not fully represented among the alleles in the mixture. The danger is that this approach may produce apparently strong evidence against a surprisingly large fraction of noncontributors. In this paper, the risk associated with the approach of ignoring loci with discordant alleles is assessed by simulation.

KEYWORDS: forensic science, dropout, inclusion probability, exclusion probability, RMNE, likelihood ratio, mixtures

In forensic DNA analysis, a profile representing two or more contributors is termed a mixed DNA profile. There are two predominant methods used for interpreting mixed DNA profiles. These are likelihood ratios (LR) and inclusion probabilities. The latter can be one of: CPI—cumulative probability of inclusion, RMNE—random man not excluded, or cumulative probability of exclusion (1), all of which are mathematically equivalent. There is a large body of literature detailing the methodology for the creation of LR (2–10), but much less detailed methodology has been offered for inclusion probabilities. The interpretation of mixed profiling results becomes markedly more challenging when the amount of DNA template is such that dropout of alleles is possible (4–9).

Recent discussions on a forensic discussion group highlighted the prevalence of a practice in the application of inclusion probabilities when dropout is possible that is of significant concern. In such cases, we are aware of an unpublished practice involving calculation of an inclusion probability only for those loci at which the profile of interest (hereafter the suspect) is fully included among the alleles present in the crime scene sample and to omit those loci at which the suspect has alleles that are not fully represented among the alleles in the mixture. For example, if the crime scene sample has alleles *a*, *b*, and *c* and the suspect is of type *ab*, then the inclusion probability is calculated for that locus, whereas if the suspect is of type *ad*, then no inclusion probability is calculated.

The danger is that this approach may produce apparently strong evidence against a surprisingly large fraction of noncontributors. There are many theoretical or philosophical approaches that highlight the risk associated with such *post hoc* analyses (see for example [10]). In this case, the *post hoc* part of the analysis is the selection of those loci for interpretation after comparison with the suspect's profile.

An allele present in a suspect that is not present in a mixed profile is termed a discordant allele. While it is possible that the allele has dropped out this is not certain and cannot be inferred from its presence in the suspect and absence in the mixed profile. Proceeding by ignoring the locus or treating it as uninterpretable is often considered conservative but there have been previous warnings that this is not always so (2,3,11). It is possible that the findings of these papers have not impacted on inclusion probability thinking because of the likelihood ratio framework adopted in them. However, the findings do indeed apply to inclusion probabilities.

In this paper, the probability of an adverse outcome associated with the approach of ignoring loci with discordant alleles is assessed by simulation.

Inclusion and Exclusion Probabilities

We define the quantities of interest in this section to aid the reader in the interpretation of the remainder of this paper. If a mixture has alleles $A_{l1}, A_{l2}, \ldots, A_{ln_l}$ at locus $l = 1, ., L$, then the probability that a random man would be included at locus l is defined as $PI_l = \left(\sum_{i=1}^{n_l} p_{A_i l}\right)^2$. This is the sum of the allele probabilities for the alleles seen at that locus squared. Therefore, the probability that a random man would be excluded at this locus is

$$PE_l = 1 - PI_l$$

The (cumulative) probability of inclusion across multiple loci (assuming linkage equilibrium) is given by the product of the inclusion probabilities at each locus.

$$CPI = \prod_{l=1}^{L} PI_l$$

CPI, therefore, is the RMNE probability—the probability that a random man would not be excluded from this mixture. It is worth noting that the probability of exclusion across multiple loci is obtained by taking one minus the product of the inclusion probabilities and cannot be obtained from the product of the exclusion probabilities.

[1]University of Auckland—Statistics, Private Bag 92019 Auckland, Auckland 1142, New Zealand.
[2]ESR—Forensics, Private Bag 92021, Auckland, New Zealand.

Received 27 April 2009; and in revised form 2 July 2009; accepted 23 July 2009.

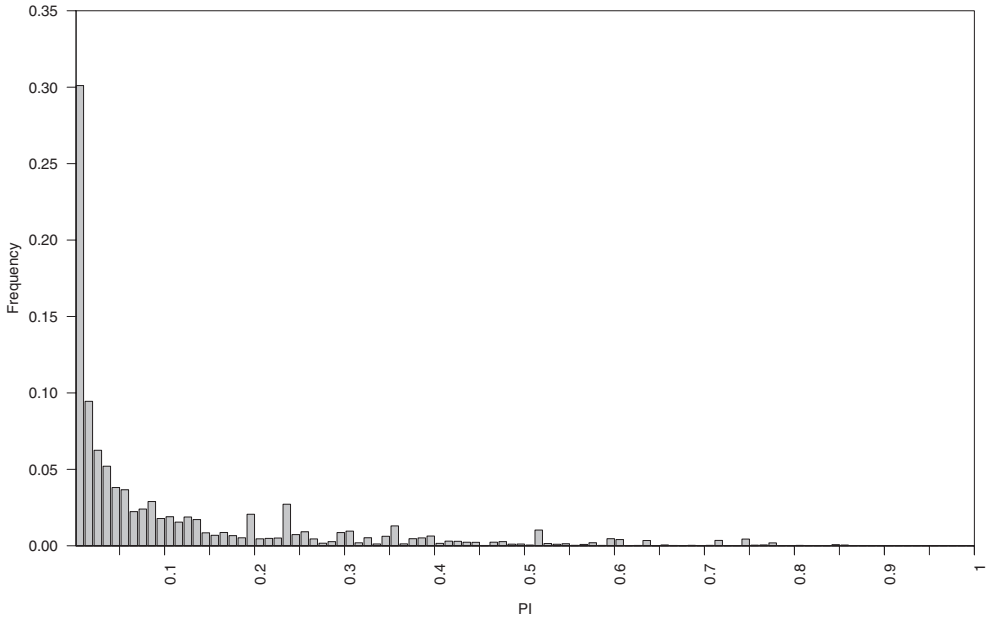


FIG. 1—*The distribution of the probability of inclusion for the 87% of cases where the evidence inculpated the random third profile. For example, 30% of included profiles had a probability of inclusion between 0.00 and 0.01.*

$$PE = 1 - CPI = 1 - \prod_{l=1}^{L} 1 - PE_l$$

and not

$$\prod_{l=1}^{L} PE_l$$

Method

There is an analytical solution to this problem. However, the solution increases in a combinatorial fashion with the addition of each locus; therefore, we have employed a simulation technique to estimate the probabilities of interest.

One thousand (1000) two-person mixtures were simulated by randomly selecting two genotypes from the CODIS loci using published African American allele probabilities (12). The inclusion probability across all 13 of the CODIS loci for these 1000 mixtures was between 5.6×10^{-11} and 3.4×10^{-5}.

A third profile having no relationship to the two contributors was randomly generated and compared with the mixture. The inclusion probability was calculated for those loci where the third profile was completely represented among the alleles of the mixture using the product rule (1). The remaining loci, where the third profile had one or two discordant alleles, were ignored. For each of the 1000 mixtures, 10,000 unrelated third profiles were created.

Results

In 87% of simulated cases, evidence was produced that had some tendency to inculpate the random third profile that was in fact not in any way a contributor to the mixture. The distribution of inclusion statistics is shown in Fig. 1. In some of these cases, the evidence was weak and may have been (correctly) disregarded. However, the risk of producing apparently strong evidence against an innocent suspect by this approach was not negligible. In 48% of cases, the inclusion statistic was less than 0.05.

Conclusion

It is concluded that the policy of calculating an inclusion statistic for mixed DNA profiles where dropout is possible by a policy of ignoring loci where the suspect is not fully represented cannot be supported. It is false to think that omitting a locus is conservative as this is only true if the locus does not have some exclusionary weight.

Acknowledgment

We warmly acknowledge the comments of two anonymous referees that have considerably improved this paper.

References

1. Buckleton J, Curran J. A discussion of the merits of random man not excluded and likelihood ratios. Forensic Sci Int Genet 2008;2(4):343–8.
2. Balding DJ, Buckleton JS. Interpreting low template DNA profiles. Forensic Sci Int Genet 2009;4(1):1–10.
3. Buckleton J, Triggs CM. Is the 2p rule always conservative? Forensic Sci Int 2006;159:206–9.
4. Curran JM, Gill P, Bill MR. Interpretation of repeat measurement DNA evidence allowing for multiple contributors and population substructure. Forensic Sci Int 2005;160:47–5.
5. Gill P, Curran J, Neumann C. Interpretation of complex DNA profiles using Tippett plots. Forensic Sci Int Genet Suppl Ser 2008;1(1): 646–8.
6. Gill P, Curran J, Neumann C, Kirkham A, Clayton T, Whitaker J, et al. Interpretation of complex DNA profiles using empirical models and a method to measure their robustness. Forensic Sci Int Genet 2008;2(2): 91–103.
7. Gill P, Kirkham A, Curran J. LoComatioN: a software tool for the analysis of low copy number DNA profiles. Forensic Sci Int 2007; 3:128–38.
8. Gill P, Puch-Solis R, Curran J. The low-template DNA (stochastic) threshold—its determination relative to risk analysis for national DNA databases. Forensic Sci Int Genet 2009;3(2):104–11.

9. Gill P, Whitaker JP, Flaxman C, Brown N, Buckleton JS. An investigation of the rigor of interpretation rules for STRs derived from less that 100 pg of DNA. Forensic Sci Int 2000;112(1):17–40.
10. Thompson WC. Painting the target around the matching profile: the Texas sharpshooter fallacy in forensic DNA interpretation. Law, Probab Risk 2009;8(3):257–76.
11. Gill P, Brenner CH, Buckleton JS, Carracedo A, Krawczak M, Mayr WR, et al. DNA commission of the International Society of Forensic Genetics: recommendations on the interpretation of mixtures. Forensic Sci Int 2006;160:90–101.
12. Budowle B, Moretti TR, Baumstark AL, Defenbaugh DA, Keys KM. Population data on the thirteen CODIS core short tandem repeat loci in African Americans, US Caucasians, Hispanics, Bahamanians, Jamaicans and Trinidadians. J Forensic Sci 1999;44:1277–86.

Additional information—reprints not available from author:
James M. Curran, M.Sc., Ph.D.
Department of Statistics
University of Auckland
P.B. 92019
Auckland
New Zealand
E-mail: curran@stat.auckland.ac.nz

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01426.x
Available online at: interscience.wiley.com

PAPER

CRIMINALISTICS

Simon J. Walsh,[1] *Ph.D.; James M. Curran,*[2] *Ph.D.; and John S. Buckleton,*[3] *Ph.D.*

Modeling Forensic DNA Database Performance*

ABSTRACT: Over the past decade or more, DNA databases have been a focal point of development for the forensic field. Using this approach, forensic and law enforcement agencies have aided millions of investigations, many of which would remain unsolved but for the intelligence links provided from DNA database comparison. However, despite their widespread use and increasingly broad legislative and operational reach, there has been limited overarching performance modeling or reflection on drivers of operational or financial efficiency. This study derives an inferential model for DNA database performance using data from major national DNA database programs. Parameters that optimize desirable database outputs (matches) are isolated and discussed, as is an approach for maximizing financial efficiency and minimizing ethical impact. This research takes important steps toward identifying measures of performance for forensic DNA database operations.

KEYWORDS: forensic science, DNA database, Combined Offender DNA Index System, U.K. NDNAD, Canada, New Zealand, intelligence, performance management

Forensic DNA databases have altered the landscape of the criminal justice system (CJS) and reshaped the field of forensic science. They have provided new challenges to the mechanisms by which forensic evidence can be utilized and have brought pressures and increased responsibility upon those who administer their use. There has been widespread review and commentary regarding the legal and socio-political basis of DNA databases, but there remains a lack of meaningful empirical assessment of database performance and effectiveness.

To assess the database performance we need to define and decide how we might measure the success of forensic DNA databases. To do this effectively is a complex undertaking that requires the coalescence of a range of experimental methodologies across numerous domains of society. Most jurisdictions do not monitor the performance of their databases beyond reporting a one-dimensional index of output relating to the number or proportion of hits. We believe that this is a major omission.

To date, there has been little attention given to the assessment of database performance—due mainly to the fact that through their entire history, there has been minimal demand for such evaluation. Forensic DNA databases have always provided results, many of which involve spectacular and unprecedented contributions to the most serious of cases. These results have taken little strategic thought to achieve, and to some degree this will always be the case. However, an era is arriving where more profound performance management and consideration of database effectiveness will be essential to ensure an ongoing contribution and to manage future challenges.

As a database ages a number of changes occur, some of which are detrimental. Consider the crime sample database. These are samples from crime scenes and may be from the true offender or may be from an irrelevant source—such as the home owner of a house that has been burgled. Scenes of crime officers (SOCOs) seek to keep the fraction of relevant samples high. However, certain types of sample, such as cigarette butts, and surprisingly blood, have a risk of irrelevancy. As time passes, crimes are solved which sometimes results in the removal of relevant samples from the database. However, case resolution does nothing for irrelevant samples, and so those samples largely remain on the database. Therefore, the fraction of irrelevant samples on the unsolved part of the crime sample database may slowly rise.

Even relevant samples that generate hits also gradually diminish in their overall value over time. This is because resolving old crimes is of less value and often harder than resolving recent crimes. While this is a general statement it is certainly true of property crimes, which make up the vast bulk of database cases and hits.

The database which is composed of samples from people also ages. In a very coarse model this database is composed of active offenders and nonoffenders. In general, there is a movement of some active offenders toward nonoffending. This occurs as active offenders retire from crime, die, or become imprisoned. The fraction of active offenders on the person database has a tendency to decline over time. This decline can be exaggerated if selection criteria for inclusion on the database decreases the fraction of offenders sampled. We acknowledge that this model is very simplistic and that the division of people into active offenders and nonoffenders overlooks the large range of behaviors exhibited by people. However, we have found this thinking very useful when considering the value of database sampling strategies and how they interrelate with performance.

[1]Forensic & Data Centres, Australian Federal Police, Canberra, ACT, 2601, Australia.
[2]Department of Statistics, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand.
[3]ESR Ltd., Private Bag 92-021, Auckland, New Zealand.

*Research emerges from the Ph.D. research of Simon J. Walsh undertaken at the University of Technology, Sydney (UTS). The project was supported by a UTS Industry Link Grant.

Received 29 Sept. 2008; and in revised form 8 July 2009; accepted 29 July 2009.

An additional factor that further compromises the efficacy of the person database is that there is a small but almost unavoidable accrual of duplicate samples as people are sampled twice due, for example, to subterfuge, sampling under aliases, or the existence of numerous legislative models feeding into the one national database (each requiring samples to be collected from locally apprehended suspects or offenders). This latter example exists in Australia.

Presently, assessment of the effectiveness of forensic DNA databases is almost exclusively limited to primary indexes such as profile inclusions, "match rates," or "hit rates (HR)." Generally, these are calculated by the total number of hits (or matches) divided by the total number of crime samples loaded to the Crime Sample (or Forensic) Database. Crime samples may come into the forensic laboratory from no-suspect cases or from cases where the police already have a suspect. There is logic to putting those cases where the police already have a suspect onto the database. This may be to see whether that crime links to any others or it may simply be that the police already know that the suspect is on the database and are conserving resources. The hits in the no-suspect type case are termed "cold" hits but in those cases where there is a suspect this term is not appropriate. Such hits are sometimes referred to as "warm" hits. Different organizations have different reporting policies to these cold and warm hits, and this leads to difficulties in comparison. This is just one factor that makes comparison difficult.

In the U.S.A., the Combined Offender DNA Index System (CODIS) system measures effectiveness in terms of "investigations aided"—an index defined as the number of cases/investigations that CODIS assisted through a hit where a hit is defined as a match produced by CODIS that otherwise would not have been developed (1,2). These indices are one-dimensional and limited in their informativeness, particularly in relation to using them to determine optimal strategies for database operation. They measure output (hits) rather than outcomes (verdicts) and are not corrected for influencing factors such as cost.

In this article we suggest some statistics that could be used to measure the effectiveness of databases (as defined largely by hits). From this modeling it is possible to construct hypotheses regarding the factors that drive or limit DNA database effectiveness.

For the purpose of this comparison, all persons' databases were referred to collectively as "offender" databases. It is appreciated that this is a crude simplification and is dismissive of the legislative preconditions in a number of jurisdictions under study; however, it simplifies the discussion for the purpose of this comparison.

This type of analysis requires comparable data. However, some organizations include the crime sample from crimes where there is a strong suspect on the crime sample database. If the suspect is also included on the suspect database, then this results in an elevated HR from warm hits. In addition, some laboratories appear to report the size of the crime sample database differently with some leaving in solved crimes and others removing them.

One outcome of this work is the awareness of the need for data to be collected in a way that allows international comparisons.

Inferential Statistics—Modeling Forensic DNA Database Performance

A summary of the nomenclature is given in Appendix 1.

We model a database containing profiles from persons sampled for forensic purposes of size N_t at time t as consisting of two categories—one of active and one of either inactive or nonoffenders. We denote the fraction of active offenders on the database at time t to be α_t. This fraction of active offenders will be responsible for some, but not all, of the samples in the crime sample database.

If people are selected completely at random for placement onto the database, then the fraction of active offenders recruited should initially be equal to the fraction of the population that is actively criminal hence

$$\alpha_t = \frac{M_t}{P}$$

where P is the overall population size and M_t is the size of the active offender population (whether sampled or not).

If people are targeted for inclusion onto the database because of perceived risk of future offending, then α_t should exceed this fraction, at least in the early life of a database, i.e.,

$$\alpha_t \geq \frac{M_t}{P}$$

Consider a crime sample database at time t of size C_t. This database will consist of relevant and irrelevant profiles. By relevant we mean profiles that have indeed come from the perpetrator, whereas irrelevant means that they have no link to the crime itself. We will call the fraction of the crime sample database at time t that is relevant ω_t. Consider the number of hits, H_t, expected when we compare the crime sample database with the person database. There are three types of hit possible which we will term relevant, H_D, irrelevant, H_I, and adventitious, H_A. The relevant hits are between a sample from the perpetrator and the true donor. Logically this can be modeled as:

$$H_D = \frac{\alpha_t N_t}{M_t}\omega_t C_t \tag{1}$$

If we regard M_t as being relatively constant over time, i.e., $M_t \approx M$, then

$$H_D \approx \frac{\alpha_t \omega_t N_t C_t}{M}$$

Irrelevant hits are modeled as

$$H_I = \frac{(1-\omega_t)N_t C_t}{P}$$

and adventitious hits are modeled as

$$H_A = \sum_{i=1}^{N}\sum_{j=1}^{C} f_{ij}$$

where f_{ij} is a suitable estimate of the match probability between profile i on the person database and sample j on the crime database.

Total hits, H_t, are the sum of these three contributions

$$H_t = H_D + H_I + H_A$$

If suitable care is taken to maintain a match criterion that keeps adventitious hits low then H_A should not be a large term, but could easily become so if a database contained partial crime profiles, historical person profiles, say, typed only at the six SGM loci, and hits were reported irrespective of quality. However, we assume that suitable care is taken and that H_A is low. Inspection of the equation for H_I suggests that this should be a small term if $(1-\omega_t)N_t/P$ is

small which will occur if the sampling of scenes effectively targets profiles from the true perpetrator or the fraction of the population sampled for the person database is small, or both. For simplicity we ignore irrelevant or adventitious hits but there is no mathematical requirement to do so. It seems likely that at least some samples from the scenes of volume crime are not from the offender and may in some cases be from the complainant. Hence, it may be wrong to think of ω_t as a number close to 1.

Therefore, we anticipate the HR_t at time t to be proportional to the fraction of the active offender population currently sampled. We model this as

$$HR_t = \frac{\alpha_t \omega_t N_t}{M_t}$$

The model, $H_D \approx \frac{\alpha_t \omega_t N_t C_t}{M}$, suggests that the number of relevant crime to person hits is determined in the early stages of database development predominantly by the product of the number of people on the person and crime databases times two quality factors (α_t, ω_t, respectively) that are measures of the appropriateness of the sample selection.

Key elements of this relationship may appear self-explanatory, for example, that increasing N_t and C_t should correspondingly increase H_t. Although it is obvious that the product $N_t C_t$ should be the most relevant measure of total database capacity, this was novel thinking to the authors. It is our view that current database models have varied in their attention to each of these factors. For example, political proponents of sampling on arrest or sampling everyone focus on increasing N_t, but not on increasing C_t. It is also anticipated that the notion of quality factors will be highly instructive and link ultimately to database performance.

Next, consider an increment of time such that the new time is $t + 1$. We consider that there is a rate of retirement from active criminal behavior whereby persons (while they remain on the database) move from the active to inactive compartments. We consider this rate to be μ per unit time. We consider the number of people n added to the N_t initially on the database. We term these people "new people" for short. The database after the addition of n people has N_{t+1} person samples. These n people will comprise people who:

- Have become active in the interval t to $t + 1$ but were not active prior to t,
- Were active prior to t and are still active in the interval t to $t + 1$,
- Were active prior to t but now are not active in the interval t to $t + 1$, and
- Were not active either before or in the interval t to $t + 1$.

Denote the fraction of new people in each of these categories as $\pi_{\bar{a}a}, \pi_{aa}, \pi_{a\bar{a}}, \pi_{\overline{a\bar{a}}}$, respectively, such that $\pi_{\bar{a}a} + \pi_{aa} + \pi_{a\bar{a}} + \pi_{\overline{a\bar{a}}} = 1$. Consider that $C_{t+1} - C_t = c$ crime samples are added to the crime sample database over this time period, termed "new crimes" for short. Of these ω are from the offender. We expect new hits from

- New people to new crimes, modeled as $n(\pi_{aa} + \pi_{\bar{a}a})\omega c$.
- Old people to new crimes, modeled as $\alpha_t N_t(1-\mu)\omega c$.
- New people to old crimes $n(\pi_{aa} + \pi_{a\bar{a}})\omega C_t$.

Hits of newly added people to old crimes are often termed "back hits." They may be of lesser value especially for volume crime but not necessarily for serious crime. We will term hits to the newly added crimes "recent crime hits."

This suggests that the new hits in the time period t to $t + 1$ should be

$$H_{t+1} - H_t = \frac{n(\pi_{aa} + \pi_{\bar{a}a})\omega c + \alpha_t N_t(1-\mu)\omega c + n(\pi_{aa} + \pi_{a\bar{a}})\omega C_t}{M_{t+1}}$$
$$= \frac{\{n(\pi_{aa} + \pi_{\bar{a}a}) + \alpha_t N_t(1-\mu)\}\omega c + n(\pi_{aa} + \pi_{a\bar{a}})\omega C_t}{M_{t+1}}$$

Maximizing new hits for a given n and c is achieved if attention is paid to ensuring high $\pi_{\bar{a}a}, \pi_{aa}, \pi_{a\bar{a}}$ and ω. High values for these terms suggest that authorities are sampling active or previously active persons and relevant crime samples. To maximize recent crime hits authorities should seek to sample actively offending persons and relevant crime samples hence maximizing $\pi_{\bar{a}a}, \pi_{aa}$ and ω. These conclusions are relatively obvious but the thinking is instructive nevertheless. Of greatest significance here is that back capture of prison inmates or older persons with prior offending stands a risk of maximizing $\pi_{a\bar{a}}$. Maximizing $\pi_{\bar{a}a}, \pi_{aa}$ tends to direct authorities to emerging and youthful offenders at least in part.

If authorities' sampling strategies are not changing markedly over time, then $\alpha_t = \alpha$ and $\pi_{\bar{a}a} + \pi_{aa} \approx \alpha$. In other words we are assuming that the fraction of currently active offenders in the newly added people, n, is about the same as it was when we first sampled the older set, N_t. With less confidence we consider that $\pi_{aa} + \pi_{a\bar{a}} \approx \alpha$ which assumes that sampling of persons who were active in the past but are sampled now is performed with the same efficiency as sampling of persons who are currently active.

$$H_{t+1} - H_t \approx \frac{\{n\alpha + \alpha N_t(1-\mu)\}\omega c + n\alpha\omega C_t}{M_{t+1}}$$
$$\approx \alpha\omega \frac{\{[n + N_t(1-\mu)]\}c + nC_t}{M_{t+1}}$$

Adding in the previous hits

$$H_{t+1} \approx \alpha\omega \frac{\{[n + N_t(1-\mu)]\}c + nC_t}{M_{t+1}} + H_t$$
$$\approx \alpha\omega \frac{\{[n + N_t(1-\mu)]\}c + nC_t}{M_{t+1}} + \frac{\alpha\omega N_t C_t}{M_t}$$

Assuming $M_{t+1} \approx M_t$

$$H_{t+1} \approx \alpha\omega \frac{N_{t+1}C_{t+1} - \mu N_t c}{M_{t+1}}$$

Comparison of this with the equation for hits at time t: $H_t = \frac{\alpha\omega N_t C_t}{M_t}$ suggests that the gradient of a plot of H_t versus $N_t C_t$ is expected to level over time given all other factors being equal as an increasing fraction of persons (μN_t) progress from the active to the inactive category. This result is intuitively reasonable and much as expected. Any changes from the expected curve in the upward direction suggest that the authorities have improved their process in one or more of the factors highlighted.

Testing the Inferential Model Using CODIS Data

We sought to investigate the relationships proposed by this model using the 52 SDIS databases that comprise the CODIS dataset (3). These data were used, as it is our most homogenous set of data. It was obtained from the FBI website over consecutive months from June 2002 to July 2007. We examined the predicted relationship between H_t and $N_t C_t$. This is shown in Fig. 1, below. Recall that this relationship is predicted to level over time.

Taken broadly, the data appear to support the suggested relationship well. It is possible even to perceive the levelling in the upward

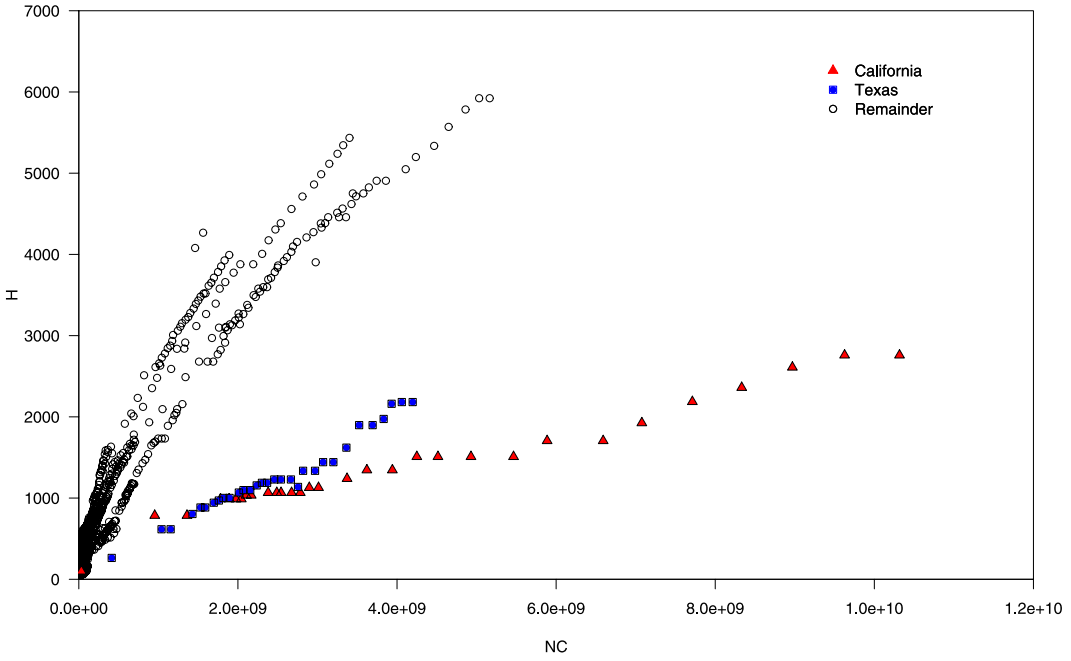


FIG. 1—*Scatter plot showing the relationship between* H *and* NC *for 50 SDIS databases of the Combined Offender DNA Index System (CODIS).*

slope predicted by the model as consecutive cohorts of active criminals retire to the inactive compartment of the database. The broad coherence of the data with the predictions of the model enhances our faith in the analysis.

Our model suggests

$$H_t \propto \frac{\alpha_t \omega_t N_t C_t}{M}$$

Simple rearrangement of this equation suggests an efficiency measure, the return index (RI_t) at time t:

$$RI_t = \frac{H_t}{N_t C_t} = \frac{\alpha_t \omega_t}{M}$$

We propose four possible causes for a low RI_t:

(a) a suboptimal ratio of C_t:N_t, or,
(b) a person sampling strategy that leads to a low recruitment of active criminals (low α_t),
(c) a crime scene sampling practice that too often obtains samples from items not associated with the crime and which lead to irrelevant profiles (low ω),
(d) high active criminal population (high M).

Hypothesis (a) can be plausibly examined by exploration of the data already described. A plot of RI_t versus C_t/N_t should show a trend if C_t/N_t has any significant influence on the efficiency metric. This plot is shown below for the full set of CODIS data (Fig. 2). It would be very difficult to see a trend in this data. We therefore conclude that there is very little support from this data for hypothesis (a).

We turn next to the examination of hypotheses (b) and (c). As discussed earlier, our putative model suggests that

$$\mathrm{HR}_t = \frac{\alpha_t \omega_t N_t}{M}$$

For each state we suggest that the active criminal population M is equivalent to some fraction δ of the total population P; hence,

$$\mathrm{HR}_t = \frac{\alpha_t \omega_t N_t}{\delta P}$$

This suggests that HR_t should be proportional to N_t/P at least initially with levelling if persons "retire" from crime.

While the outcomes of this comparison (shown in Fig. 3) are not as stark as we may have predicted from the model some general trends are observable that do concur with our predictions. First, the databases with the highest HR also tend to be those databases that have surveyed a higher fraction of the population. Second, the majority of the data (in the remainder category) shows a trend in accordance with that predicted, that is increasing N_t/P correlates weakly to increasing HR. The trendline on this plot provides a rudimentary distinction between datasets with effective or ineffective recruitment practices, crime sampling practices or those states which by good providence or good management happen to have low active offender populations. Put another way, databases below the trendline (to the right) are as follows:

- either recruiting a low proportion of active offenders (low α), or
- are sampling inefficiently at crime scenes (low ω), or
- they have a high offender population.

Applying the Inferential Model to Other National DNA Databases

The next obvious step was to take the model to disparate datasets. We suggest that this model has the potential to inform thinking on the performance management and improvement of DNA databases. By utilizing the model to benchmark datasets in the RI continuum and by focussing the attention of relevant agencies of the variables under their control (α and ω), we envisage a significant potential for rational strategies for improvement that are founded in local performance data. As test datasets, the United Kingdom (U.K.), New Zealand (NZ), and Canadian data have been used.

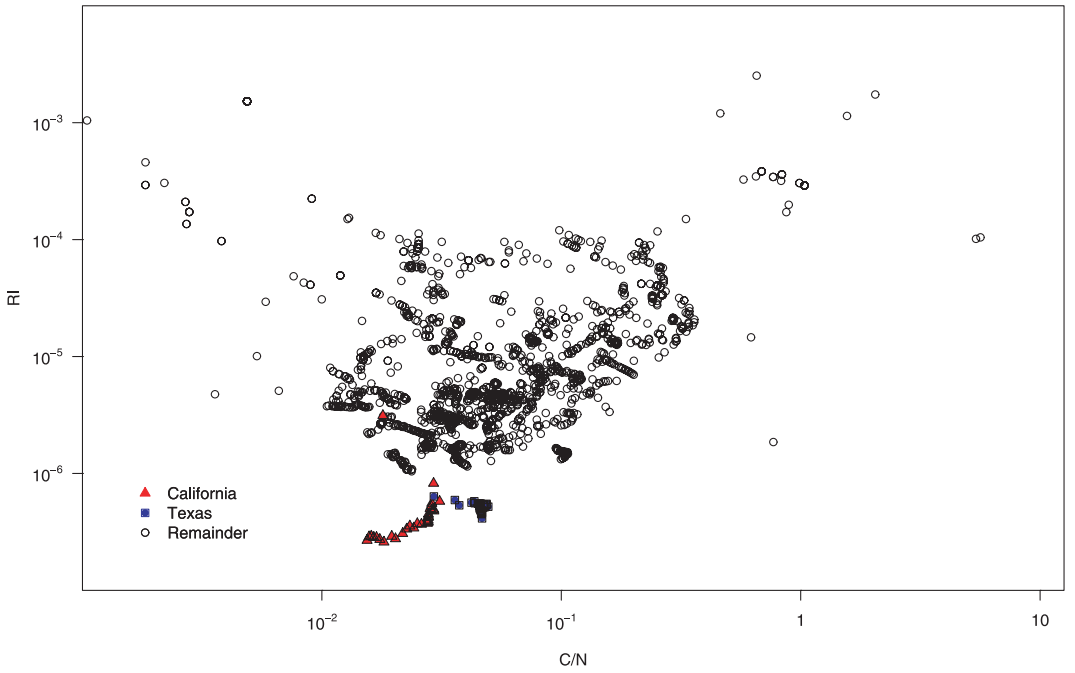


FIG. 2—*Scatter plot showing the relationship between RI and the ratio of* C:N *(log scale) for 50 SDIS databases of the Combined Offender DNA Index System (CODIS).*

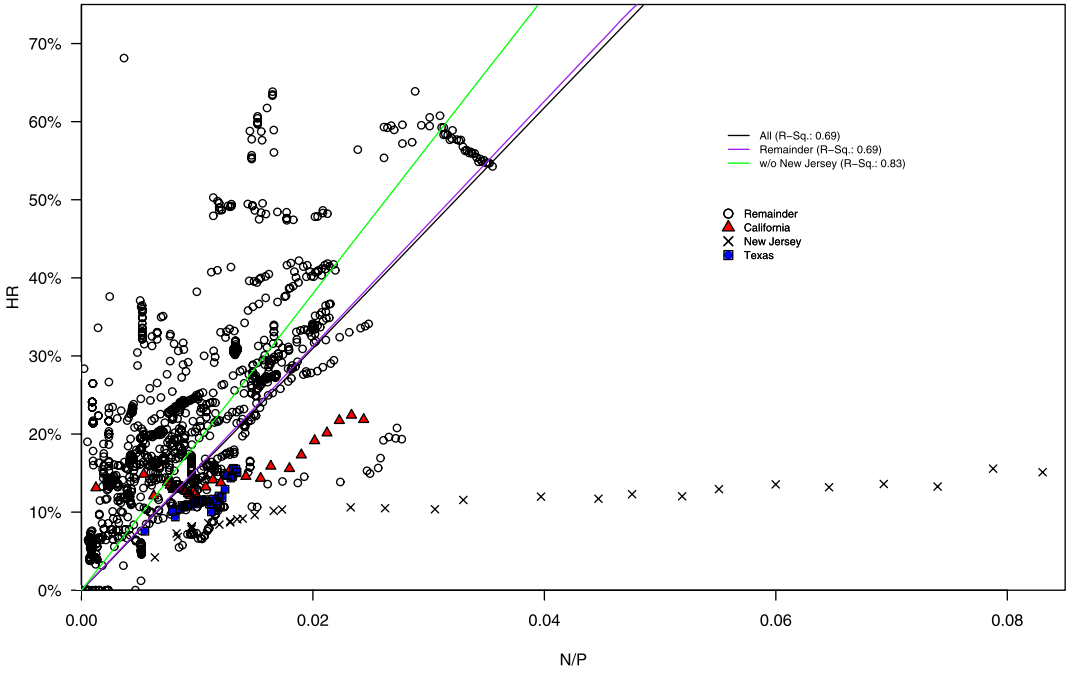


FIG. 3—*Scatter plot showing the relationship between hit rate (HR) and the proportion of the population sampled* (N/P) *for 49 SDIS databases of the Combined Offender DNA Index System (CODIS). The FBI/DC Laboratory data were removed from this comparison as it is not expected that the database is populated only by persons and crimes from the District of Columbia.*

United Kingdom (U.K.)

The U.K. National DNA Database (NDNAD) is the most frequently discussed NDNAD model. It is the oldest national database system and has been until very recently consistently the largest. As a result, most forensic DNA database operations benchmark themselves against the U.K. model and its performance. For this component of the research, public domain data describing the size and outcomes of the U.K. NDNAD were utilized (4).

A plot of H_t versus N_tC_t is expected to plateau as the database ages and successive cohorts of offenders move from active to inactive. This plot for the U.K. NDNAD is shown in Fig. 4 (below). The observations fit the predictions of the model very well.

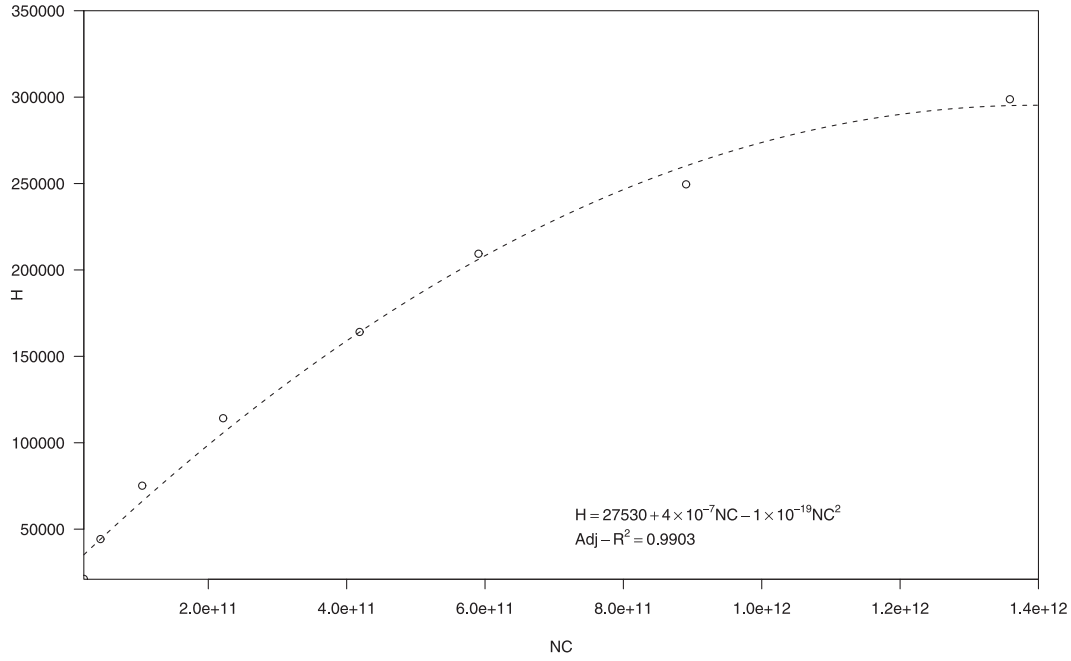


FIG. 4—*Scatter plot showing the RI relationship (*H/NC*) for the U.K. National DNA Database.*

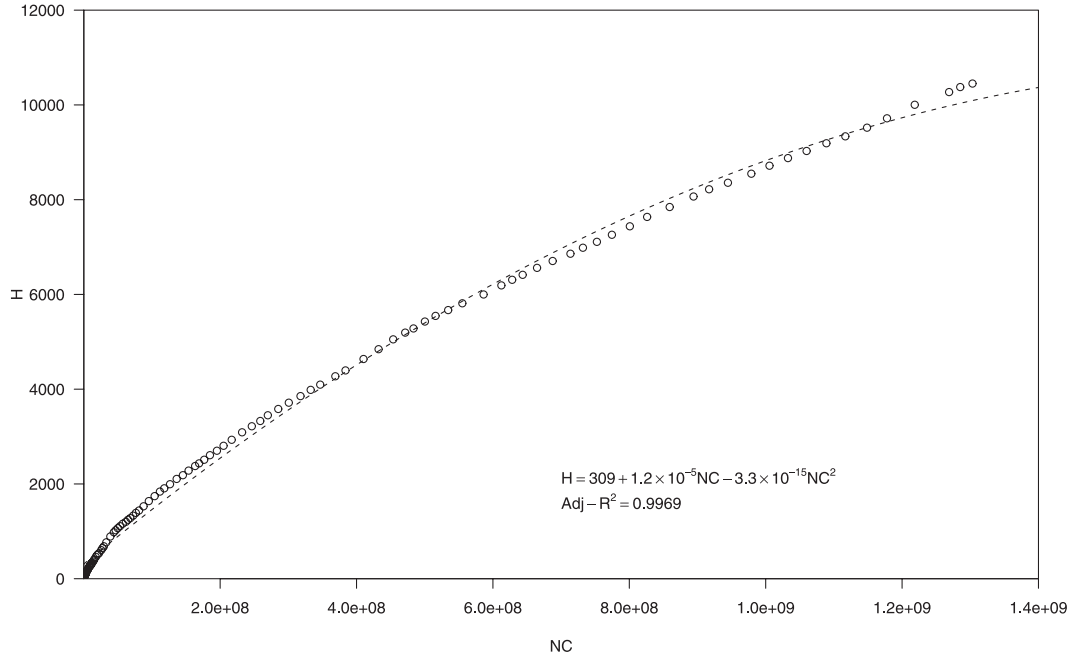


FIG. 5—*Scatter plot showing the RI relationship (*H/NC*) for the NZ National DNA Database.*

New Zealand (NZ)

The NZ NDNAD began operations in 1996, 2 years after the U.K. NDNAD, and was the second national system implemented. Data were obtained directly from the Institute of Environmental Science and Research Ltd. In global terms, the database remains small but it contains over 2.1% of the NZ population, which is a higher proportion than the other countries analyzed herein (U.S.A. 1.7% and Canada 0.5%) with the exception of the United Kingdom (7.0%). Figure 5 shows the H_t versus N_tC_t relationship for the NZ NDNAD. Again, the observations fit the predictions of the model very well.

Canada

The Canadian NDNAD began operations in January 2000. Figure 6 shows the relationship of H_t versus N_tC_t for the Canadian NDNAD, which is very close to linear. Data were obtained from the official website administered by the Royal Canadian Mounted Police (5).

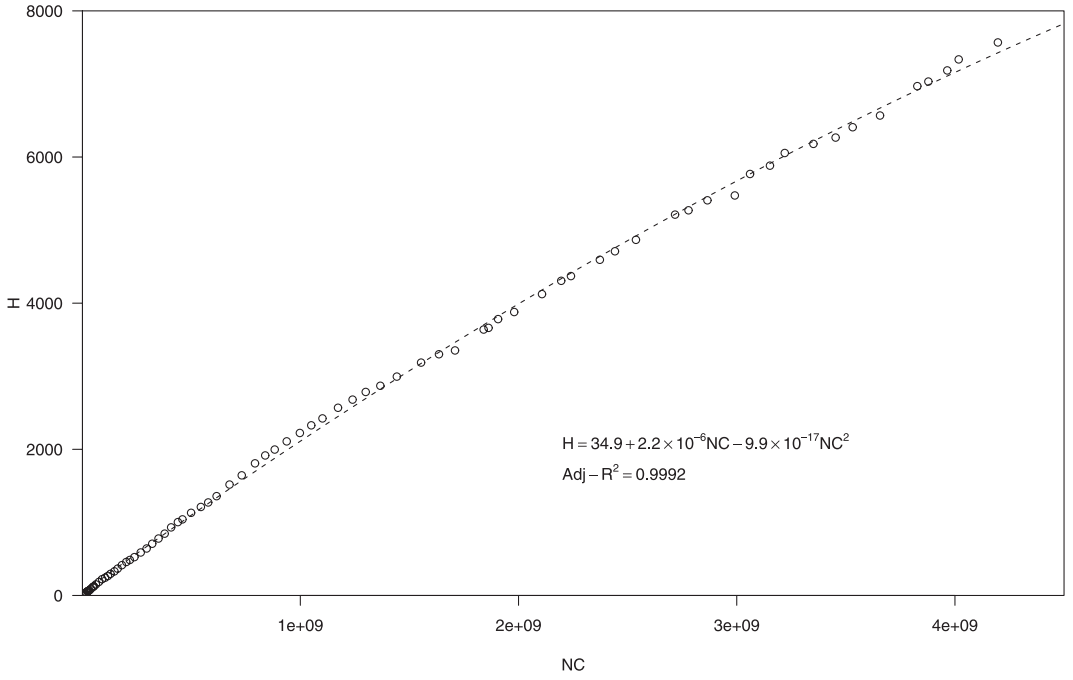


FIG. 6—*Scatter plot showing the RI relationship (*H/NC*) for the Canadian National DNA Database.*

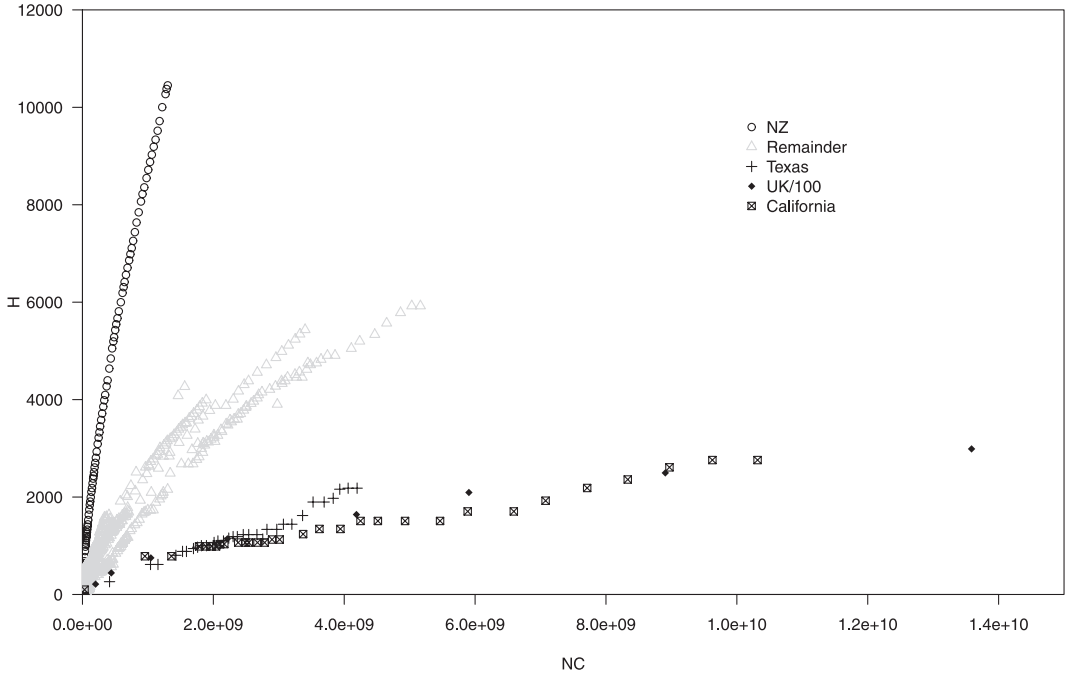


FIG. 7—*Scatter plot showing a comparison of the RI relationships (*H/NC*) for the Combined Offender DNA Index System (CODIS), NZ, and Canadian data against a background of the CODIS data.*

Combined

The combined data are shown in Fig. 7. We acknowledge that different recording practices significantly affect this analysis but are beyond our reach.

In this figure, the U.K. NDNAD data have been included but both *H* versus *NC* required division by a factor of 100 to allow the data to be visible on a single plot. When the combined data are examined, it is possible to see the general pattern of plateauing in much of that data. This also allows an absolute comparison. The inferential model suggests that the most effective measure of database size is the product N_tC_t. This is good news for database operators as this product rises more rapidly than either N_t or C_t separately. This suggests that there is a driver for increasing return as the database grows. This is offset by retirement from the actively offending portion of the database. However, the use of the product N_tC_t allows us to relate databases of different absolute sizes and different ratios of N_t to C_t. Any database that has a high

number of hits (H_t) for its size (N_tC_t) is operating well. Our model suggests that this may occur because of the good selection of suspect samples (high α_t), good scene work and sample selection (high ω), or a low active criminal population (low M).

The RI is our best attempt at a size independent measure of return. It is not quite size independent because large databases are usually older. Hence, size, age, and retirement from the active criminal element are confounded. However, we see RI_t as a useful measure of return. Any database with a high RI_t, irrespective of size, is being well operated and any change in policy that improves RI_t is likely to improve effectiveness.

Factoring in the Financial and Ethical Cost

The analysis to date has identified some useful measures for benchmarking performance and assessing policy changes but has not directly encompassed the financial and ethical elements of databasing.

It is evident that crime samples are more financially costly to process and include upon a database than suspect or other person samples. This is attributed to the increased complexity of handling and interpretation that in turn reduces the applicability of automated or expert platforms. In contrast, crime samples are more ethically acceptable to collect than suspect or person samples (Table 1). In this section a discussion and analysis is presented that attempts to weigh or incorporate cost.

Financial Cost

A coarse financial analysis based on the inferential model derived earlier is attempted. Assuming that processing a crime sample costs x times as much as processing a suspect or person sample suggests that total cost, Z_t, is proportional to $N_t + xC_t$ where $x > 1$.

$$Z_t \propto N_t + xC_t$$

The return on investment may be modeled as the hits H_t divided by the cost. This suggests

$$\frac{H_t}{Z_t} \propto \frac{N_tC_t}{N_t + xC_t}$$

Assume that the task is to find the optimal mix of n and c that maximizes return for a fixed cost, Z_1. Recall that $Z_t = k(N_t + xC_t)$; hence

$$N_t = \frac{Z_1}{k} - xC_t$$

To obtain the best return it is necessary to maximize N_tC_t subject to the constraint that $k(N_t + xC_t) = Z_1$ where Z_1 and k are constants. The maximum return on investment is $\frac{Z_1^2}{4k^2x}$ and occurs when $\frac{N_t}{C_t} = x$ (see Appendix 2). This implies that return on investment increases quadratically with increased investment, Z_t.

In summary, to optimize the return per unit investment, database administrators should set the ratio of N_t to C_t so that it is equal to the cost ratio of C_t to N_t. If this ratio is achieved, then return per unit investment increases with investment.

TABLE 1—*Representation of the relative financial and ethical cost of crime and person samples.*

	Financial Cost	
Ethical Cost	Low	High
Low	–	Crime samples
High	Person samples	–

Ethical Cost

The expectation of success for any DNA database revolves around the belief that the individuals included on the database are likely to commit further crimes, and that a worthwhile proportion of unsolved crime scenes will yield DNA evidence of sufficient quality to allow analysis and comparison and the profile of the perpetrator(s). In terms of the construction of DNA databases, debate has circled around which crimes qualify as collection offenses and at what point in the process the samples should be collected. We have, unsurprisingly, identified earlier that the fraction of sampled persons who are offending is a driver for effective database operations.

Early DNA databases were often constructed using the profiles of convicted sex offenders (6). There was neither little objection, nor apparent reason to object to the rationale and intent of this sampling regime.

Throughout the CJS, one can observe that some types of offenders appear more likely to be returned to prison than others. It is important that criminological understandings form some part of what should be a considered strategy that incorporates forensic DNA databases as part of crime reduction initiatives. This philosophical standpoint is supported not only on the basis that it provides a justification to counteract the ethical imposition of mandatory sampling upon an individual or a community, but also on the basis of the empirical modeling undertaken in earlier sections which highlighted the value of thoughtful sampling strategies (high α).

In the early stages of DNA database development, therefore, a situation existed where the law enforcement community had a group of individuals that had little social or legal right to object to mandatory sampling of their DNA, and for whom it was thought the high recidivist tendencies would prevent and resolve future heinous crimes. Kaye and Smith (7) refer to these twin justifications as the *forfeiture* and *predictivist* theories. Pressure soon came to extend the reach of DNA database sampling beyond this core category of entrant. Although empirical data to support the predictivist justification were unconvincing, DNA technology was beginning to prove itself capable of a role in a wide variety of crimes and evidence types (8–15). The result was a proliferation of legislative schemes broadening those who can be included in a DNA database, the widest inclusionary rule being inclusion upon arrest regardless of later conviction or acquittal.

There are some significant ethical and practical ramifications of such a framework. Based on racial distribution statistics from the US CJS, Kaye and Smith ([7], p. 456) fear that an arrest-only database "*would have the look and feel of a universal DNA database for black males, whose already jaundiced view of law enforcement's legitimacy is itself a threat to public safety.*" Goode (16) queries this developmental course from a more pragmatic perspective. He sees that there appears to be two alternative choices: to use DNA under special circumstances for serious offenses, or to apply it as a routine procedure in as many police investigations as possible. He argues, on the one hand, that the latter approach of "*having a huge national DNA offenders databank including every common or garden assault in or near a drinking establishment uses a sledge hammer to crack a nut and is an invasion of privacy quite out of proportion to the offence committed*" ([16], p. 70). The earlier analysis through inferential modeling lends some weight to this view in that larger sampling

regimes, over time, generate high redundancy and subsequently, a levelling or diminishing of the RI.

Some authors have discussed the intertwined considerations of *privacy* and *equality* as underpinning features of fair and effective DNA database legislation (7,17). In taking this line of thought to its extreme, Kaye (17) makes a case for a population-wide database. He suggests that an identification system based on noncoding loci (and thereby unable to reveal personal or medical information about its donors) is not a substantive invasion of privacy. However, in contrast, he feels that limiting DNA collection regimes to convicted offenders or arrestees extends the racial inequality of the CJS and diminishes the effectiveness of the system, despite being more palatable in terms of its privacy intrusions. For these reasons Kaye and Smith (7) favor a population-wide database as the best means to advance public safety, ensure racial even-handedness in the CJS and act as a firewall against far greater privacy intrusions on behalf of law enforcement agencies, such as speculative searches through medical and personal records. While this is more ethically pristine, there are considerable practical and financial barriers to implementing such an approach.

Balancing the ethical costs against maximal investigative benefit is a fundamental corollary of DNA database systems, and in fact many other law enforcement data and intelligence gathering initiatives. While it is impossible to factor ethical cost directly into our modeling there is still a plausible relationship. A key question is "What is the risk or personal cost to a nonoffender from being on a database?" The primary risk of being falsely accused arises from an adventitious match or a scientific error. These are minimized by good scene and laboratory work. The primary personal costs arising from inclusion of nonoffenders on a database would occur if such inclusion was ever used when considering employment, insurance, or similar matters, the intangible costs to a person's self-image by inclusion, and potential anger, resentment of, or loss of faith in the authorities. Some, but not all of these, are reduced by strict and enforced rules about what information is available or may be requested by potential employers or insurers.

It is not obvious whether the ethical cost increases or decreases as the sampling regime broadens. Clearly, universal inclusion actually reduces any stigma associated with the database. Universal sampling would not necessarily deliver the maximal return especially if resources were diverted from sampling crimes. Strategic sampling, which we have observed as a driver of high RI, may also be perceived as ethically compromised on the basis that it often reflects inherent biases in criminal justice representation.

Summary

We have developed and tested a simple model for database performance. We acknowledge significant limitations in this model and also acknowledge the effect that different recording practices have on this analysis. However, the predictions of the model hold particularly well when assessed on the training set (in this case the SDIS data from the FBI's CODIS system). The model provides a mechanism to cross-compare the performance of database systems operating under different legislative and operational frameworks because it is data independent. This comparison was undertaken on publicly available data from four major NDNAD programs.

The modeling suggests several key observations related to database performance management. Database performance can be assessed by a metric which we have described as the RI_t which is estimated by H_t/N_tC_t where H_t is the number of crime-to-person hits, N_t is the number of person samples on the database, and C_t is the number of crime samples on the crime sample database. A high RI_t indicates a database with a high return; that is maximum hits per person and crime sample tested.

The primary output from a database, hits (H_t), is modeled by $H_t = \alpha\omega N_t C_t/M$. This model suggests that $H_t \propto N_t C_t$ but is conditioned by two parameters α and ω, quality factors for the selection of person and crime samples, respectively. We accept that scientists have no direct control over α and ω but they can be manipulated through selection of scene samples and those persons in the database. High α and ω maximize H_t and in turn RI_t and are therefore of critical importance in optimizing database performance. This has important implications for continuing research into drivers of success across all stages of police, forensic, and database operations.

Financial efficiency is maximized when the ratio of database samples (N_t) to crime samples (C_t) is equal to the cost ratio of C_t to N_t. For example, if a crime scene sample costs \$50 and a database sample costs \$10, then the ratio for optimal return is five database samples for every one crime sample. If this ratio is achieved, then return per unit investment increases quadratically with every dollar invested.

References

1. Asplen CH, Lane SA. Considerations in the development of DNA databases. Interfaces 2003;36:1–2.
2. Loftus P. The FBI's combined DNA index system helps corrections and law enforcement professionals monitor inmates and solve crimes. Corrections Today 1999;61:68–71.
3. Federal Bureau of Investigation CODIS–NDIS Statistics. 2007 [updated 2007; cited October 17, 2007]; http://www.fbi.gov/hq/lab/codishttp://www.fbi.gov/hq/lab/codis.
4. The United Kingdom National DNA Database Annual Report 2005–2006. 2007 [updated 2007; cited August 8, 2007]; http://www.homeoffice.gov.uk/documents/DNA-report2005-06.pdf?view=Binary.
5. National DNA Databank of Canada Statistics. 2007 [updated 2007; cited]; http://www.nddb-bndg.org/stats_e.htm.
6. Murch RS, Budowle B. Are developments in forensic applications of DNA technology consistent with privacy protections? In: Rothstein MA, editor. Genetic secrets: protecting privacy and confidentiality in the genetic era. New Haven, CT: Yale University Press; 1997;212–30.
7. Kaye DH, Smith ME. DNA identification databases: legality, legitimacy, and the case for population-wide coverage. Wis L Rev 2003;3:413–59.
8. Abaz J, Walsh SJ, Curran JM, Moss DS, Cullen JR, Bright J, et al. Comparison of variables affecting the recovery of DNA from common drinking containers. Forensic Sci Int 2002;126(3):233–40.
9. Bright J, Petricevic SF. Recovery of trace DNA and its application to DNA profiling of shoe insoles. Forensic Sci Int 2004;145:7–12.
10. Harbison SA, Petricevic SF, Vintiner SK. The persistence of DNA under fingernails following submersion in water. Int Congr Ser 2003;1239:809–13.
11. Lorente M, Entrala C, Lorente J, Alvarez JC, Villanueva E, Budowle B. Dandruff as a potential source of DNA in forensic casework. J Forensic Sci 1998;43:648–56.
12. Sweet D, Hildebrand D. Saliva from cheese bite yields DNA profile of a burglar: a case report. Int J Legal Med 1999;112(3):201–3.
13. Van Oorschot RA, Jones M. DNA fingerprints from fingerprints. Nature 1997;387(6635):767.
14. Webb LG, Egan SE, Turbett GR. Recovery of DNA for forensic analysis from lip cosmetics. J Forensic Sci 2001;46(6):1474–9.
15. Wickenheiser RA. Trace DNA: a review, discussion of theory, and application of the transfer of trace quantities of DNA through skin contact. J Forensic Sci 2002;47(3):442–50.
16. Goode M. Some observations on evidence of DNA frequency. Adelaide Law Rev 2002;23:45–77.
17. Kaye DH. Two fallacies about DNA databanks for law enforcement. Brooklyn Law Rev 2002;67:179–206.
18. Lagrange Multipliers. Wikimedia Foundation. 2009 [updated 2009; cited February 18, 2009]; http://en.wikipedia.org/w/index.php?title=Lagrange_multipliers&oldid=270803132.

Additional information and reprint requests:
Simon J. Walsh, Ph.D.
Forensic & Data Centres
Australian Federal Police
GPO Box 401
Canberra, ACT, 2601
Australia
E-mail: simonjoseph.walsh@afp.gov.au

Appendix 1

Symbol	Definition
α	The long-run average fraction of the database of individuals active at the time of sampling
α_t	The fraction at time t of the database of individuals that is active
C_t	The size at time t of the database of samples from crime scenes
H_A	Adventitious hits, those between a sample from a crime scene whether relevant or not, and an individual who is not the donor
H_D	Relevant hits, those between a sample from a perpetrator and the true donor
H_I	Irrelevant hits, those between an irrelevant (non-perpetrator) sample at a scene and the true donor of this irrelevant sample
H_t	The number of hits from the start of the database to time t
HR_t	The hit rate at time t, the chance that a given crime sample will match an individual on the database
M	The long-term average of the fraction of the population that is actively criminal
M_t	The fraction of the population that is actively criminal at time t
N_t	The size of the database of samples from individuals at time t
P	The size of the population of an area, country, or jurisdiction
$\pi_{\bar{a}a}$	The fraction of the new individuals, n, who have become active in the interval t to $t + 1$ but were not active prior to t
RI_t	A function proposed to represent the return from the database at time t
$\pi_{\overline{a\bar{a}}}$	The fraction of the new individuals, n, who were not active either before or in the interval t to $t + 1$
$\pi_{a\bar{a}}$	The fraction of the new individuals, n, who were active before t but are inactive in the interval t to $t + 1$
π_{aa}	The fraction of the new individuals, n, who were active prior to t and are still active in the interval t to $t + 1$
ω	The long-term average of the fraction of the database of crime samples that is actually from the offender
ω_t	The fraction of the database of crime samples at time t that is actually from the offender
Z_t	The cost of samples in the crime and individual databases at time t

Appendix 2

The aim of this appendix is to show how to maximize N_tC_t subject to the constraint that $k(N_t + xC_t) = Z_1$ where Z_1 and k are constants. The solution to this problem is the "best return per unit investment." The method used to maximize this function is called "the method of Langrage multipliers" after the mathematician Joseph Louis Langrange. Briefly, the method works by constructing a special function (called the Lagrangian) which has no constraints on its optimal value and maximizing/minimizing that function. The optimal value is also optimal (maximum or minimum) for the original function of interest because of the way the Lagrangian is constructed. The reader is referred to (18) for a more detailed discussion. The Lagrangian for this problem is

$$f^*(N_t, C_t) = N_tC_t - \lambda(k(N_t + xC_t) - Z_1) \tag{2}$$

To find the maximum, first, Eq. (2) is differentiated with respect to C, N, and λ

$$\frac{\partial f^*(N_t, C_t)}{\partial N_t} = C_t - \lambda k$$

$$\frac{\partial f^*(N_t, C_t)}{\partial C_t} = N_t - \lambda kx$$

$$\frac{\partial f^*(N_t, C_t)}{\partial \lambda} = -k(N_t + xC_t) + Z_1$$

Second, these equations are set to zero and solved for C_t, N_t, and λ to obtain

$$C_t = \lambda k \tag{3}$$

$$N_t = \lambda kx \tag{4}$$

$$-k(\lambda kx + x\lambda k) + Z_1 = 0 \;\Leftrightarrow\; \lambda = \frac{Z_1}{2xk^2} \tag{5}$$

Setting the partial derivatives to zero and solving suggests that the maximum return per unit investment occurs along the line where $N_t = xC_t$. This can be seen by the observation that the optimal C_t is λk Eq. (3) and the optimal N_t is λkx Eq. (4); hence, the optimal $\frac{N_t}{C_t}$ is x (the factor by which the financial cost of C_t is greater than the cost of N_t).

Finally, substituting the maximum is found by substituting these solutions back into Eq. (2) which gives

$$\frac{Z_1^2}{4k^2x} \tag{6}$$

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01477.x
Available online at: interscience.wiley.com

PAPER

CRIMINALISTICS

Deborah Polanskey,[1] *B.S.; Bobi K. Den Hartog,*[2] *Ph.D.; John W. Elling,*[2] *Ph.D.; Constance L. Fisher,*[1] *Ph.D.; Russell B. Kepler*[2]*; and Bruce Budowle,*[3] *Ph.D.*

Comparison of Mitotyper Rules and Phylogenetic-based mtDNA Nomenclature Systems

ABSTRACT: A consistent nomenclature scheme is necessary to characterize a forensic mitochondrial DNA (mtDNA) haplotype. A standard nomenclature, called the Mitotyper Rules™, has been developed that applies typing rules in a hierarchical manner reflecting the forensic practitioner's nomenclature preferences. In this work, an empirical comparison between the revised hierarchical nomenclature rules and the phylogenetic approach to mtDNA type description has been conducted on 5173 samples from the phylogenetically typed European Mitochondrial DNA Population database (EMPOP) to identify the degree and significance of any differences. The comparison of the original EMPOP types and the results of retyping these sequences using the Mitotyper Rules demonstrates a high degree of concordance between the two alignment schemes. Differences in types resulted mainly because the Mitotyper Rules selected an alignment with the fewest number of differences compared with the rCRS. In addition, several identical regions were described in more than one way in the EMPOP dataset, demonstrating a limitation of a solely phylogenetic approach in that it may not consistently type nonhaplogroup-specific sites. Using a rule-based approach, commonly occurring as well as private variants are subjected to the same rules for naming, which is particularly advantageous when typing partial sequence data.

KEYWORDS: forensic science, nomenclature, mitochondrial DNA, hierarchical bifurcating approach, phylogenetics, SWGDAM, EMPOP

An alignment strategy is required to be able to search and compare mitochondrial DNA (mtDNA) sequences contained within a reference database. The current forensic mtDNA analysis is based on sequencing at least the hypervariable regions of the noncoding portion of the mtDNA genome. These approximately 600 bases of sequence are highly informative for differentiating individuals. However, it is difficult to convey the results simply as a string of bases. Instead, nomenclature strategies have been developed that identify a limited number of select bases as opposed to recording all bases in these regions. These limited bases are listed as differences with respect to the standard mitochondrial DNA reference sequence, the revised Cambridge Reference Sequence (rCRS) (1). It is essential that the alignment and resulting naming of variant sites be consistent for searching a population database for statistical inferences of the rarity of an evidentiary mtDNA sequence (2). The complication confronting any nomenclature scheme is to consistently select among multiple possible alignments that can occur for some mtDNA sequences.

Two nomenclature approaches have been proposed to describe mtDNA variation: a rule-based hierarchical approach and a system based on phylogenetic relationships. Wilson et al. (3) proposed a hierarchical approach comprised of a set of rules for consistently selecting one alignment over other possible alignments. The main criterion of the Wilson et al.'s approach is parsimony, the selection of an alignment that has the fewest number of differences compared with the rCRS. If there are still multiple alignments with an equal number of differences, additional rules are executed to select a preferred alignment. The principles of a hierarchical approach have been endorsed by the Scientific Working Group on DNA Analysis Methods (SWGDAM) and incorporated into the manually typed SWGDAM mtDNA population database (4,5). However, the Wilson et al.'s (3) recommendations were not strictly followed, which is a point that Bandelt and Parson (5) did not consider in their analysis of the robustness of the SWGDAM database nomenclature. Primarily, the historical manually derived alignments did not abide by the rule favoring insertions and deletions (indels) over substitutions.

A more intuitive hierarchical nomenclature system has been developed (6) that still is based on the parsimony criterion and has several constraint criteria but favors substitutions over indels. Within the SWGDAM database, 99.92% of the haplotypes were consistent with this revised hierarchical approach. The remaining 0.08% of sequence differences was converted to be consistent with this more effective nomenclature approach (7). The hierarchical approach evaluates regions of a sample sequence that are considered polymorphic regions. A polymorphic region is defined as an independent section of sequence that, when aligned with the rCRS, exhibits differences with respect to the rCRS (i.e., a region is a section or site where at least one individual carries a variation with respect to the entire data set). These regions may be 2–3 bases in length or longer to include neighboring bases that will provide contextual information for generating alignments. In each polymorphic region, every possible alignment between the rCRS and the sample sequence is created (within context of the nonpolymorphic regions).

[1]FBI Laboratory, 2501 Investigation Parkway, Quantico, VA 22135.
[2]Mitotech LLC, 590 Monte Alto, Santa Fe, NM 87501.
[3]University of North Texas, HSC, 3500 Camp Bowie Blvd., Fort Worth, TX 76107.

Received 8 Jan. 2009; and in revised form 12 May 2009; accepted 31 July 2009.

This set of possible alignments for each region is processed by applying hierarchical rules, and each region is named independent of all other polymorphic regions. Thus, the naming of a variant in one region does not rely on the naming of a variant residing in any other region. The final alignments of the polymorphic regions are concatenated to generate the final haplotype.

A rule-based alignment scheme is very desirable for characterizing short fragments derived from highly degraded samples such as bone and teeth or telogen hairs, which are commonly analyzed in forensic mtDNA casework. In addition, the hierarchical approach is robust because it provides for a stable nomenclature that will not change as new mtDNA types are discovered (6). New rules can be created to address novel polymorphic regions without reanalysis of previously typed sequences in the dataset.

In contrast to the hierarchical approach, an evolutionary phylogenetic approach has been proffered (5). Some limitations of this phylogenetic approach are: the nomenclature is not stable in that novel mtDNA types will require a reanalysis of previously typed samples in a dataset, not all variant sites are addressed (only phylogenetic signatures), and the routine user is required to have substantial intimate knowledge of phylogenetic relationships of mtDNA, which is not operationally practical or likely (6). Variants that occur in one to only a few individuals may not correlate with those sites that define specific haplogroups, because either the sampling is too small or they are homoplastic. For these variants, there are no consistent phylogenetic rules applied for nomenclature, and variation in nomenclature can occur (see examples below).

Forensic scientists, unlike evolutionary biologists, require exact matching sequences when searching a database, regardless of the evolutionary events that generated the mtDNA sequence of interest. The merits of each nomenclature system can be debated, but nomenclature consistency should be one of the primary criteria for a database construct. In this work, an empirical comparison between the revised hierarchical nomenclature rules and the phylogenetic approach to mtDNA type description has been conducted on the same data set to identify the degree and significance of any differences. To this end, the 5173 phylogenetically aligned haplotypes contained within the European Mitochondrial DNA Population database (EMPOP) (8) were realigned using the revised hierarchical nomenclature system (Mitotyper Rules) implemented in the Mitotyper software and the results compared.

Results

Only 51 of the 5173 samples, representing 23 different Mitotyper regions, differed between the two systems, most at a single polymorphic region (Table 1). There were 44278 polymorphic regions in the EMPOP data set as determined by Mitotyper. The overall difference at these 44278 regions between the EMPOP types and the Mitotyper Rules is only 0.12%. As expected, most of the differences between the two nomenclatures resided within the homopolymeric and dinucleotide repeat regions located at np 16180–16199 (9 regions observed in 14 samples), np 300–318 (6 regions observed in 11 samples), np 513–526 (2 regions in 7 samples), and np 568–582 (2 regions in 5 samples). These differences are primarily attributable to Mitotyper selection of an alignment with the least number of differences with respect to the rCRS (Table 2).

In the EMPOP types, there were a few examples of inconsistent naming of identical regions. For instance, there was a difference at np16180–16199 in which samples USA0600921, SVN0600003, ITA0500260, and USA0601079 were described one way and sample POL0600375 was named differently (Table 3). Neither type corresponded to the Mitotyper result. In the np 300–318 region, there were two EMPOP-named sequences that were named differently; SVN0600092 and HUN0500202, the latter being consistent with Mitotyper (Table 3).

TABLE 1—*Results of types in EMPOP versus types generated by Mitotyper™.*

	# of Samples with EMPOP Types Different from Mitotyper Types	# of Regions with EMPOP Types Different from Mitotyper Types	# of Regions with more than One EMPOP Type for the Same mtDNA Region
rCRS range (np)			
16180–16199	14	9	1
37–51	1	1	1
57–62	9	1	1
242–255	3	1	0
300–318	11	6	1
455–461	2	1	0
513–526	7	2	0
568–582	5	2	0
Totals	52	23	4

KEN0500056 is counted twice because it has differences in two regions. The total number of unique samples shown is 51. EMPOP, European Mitochondrial DNA Population database; rCRS, revised Cambridge Reference Sequence.

Outside the homopolymeric regions, several samples aligned by Mitotyper were consistently discordant with the EMPOP phylogenetic approach. Such differences in EMPOP (0.06%) occur because these nucleotide changes are variants that reside primarily at regions that do not define haplogroups (private variants) (Table 3). For example, sample POL0600267 was typed with a 42.1C alignment, and samples HUN0600207, AUT0500255, KEN0500083, KEN0500041, KEN0500073, and USA0600739 were typed as 44.1C by the EMPOP approach (the 44.1C type is consistent with the Mitotyper result). Thus, if a forensic evidence sequence of approximately 100 bases in length was typed as a 42.1C, only one match would be observed, even though there should be a total of seven matching sequences. The same would hold true for the 57.1C (GRC0500319 and GRC0500123) and 58C, and 60.1T (HUN0500135, GRC0500023, GRC0500024, HUN0500250, HUN0500193, ITA0500300, GRC0500047, KEN0500066, and HUN0500045) EMPOP typings where only two matches would be found if the more parsimonious 57.1C type were used (consistent with Mitotyper). The Mitotyper software does not suffer the potential for nomenclature inconsistency at private mutations, because it names with a set of rules all variants, not just phylogenetic signature ones.

Conclusions

This study shows that the two nomenclature approaches, the hierarchical Mitotyper Rules and the phylogenetic approach used in EMPOP, show a high degree of concordance. The two systems generally name haplotypes in the same manner even though they use different strategies. In the majority of cases, searching a reference database for matching sequences with a type derived from either nomenclature will not result in missed sequences that could have aligned similarly. However, because the EMPOP phylogenetic approach does not have formal rules for typing private variant sites, the possibility of inconsistent alignments for concordant regions exists. This inconsistency may exacerbate searching a database for concordant types with a partial sequence derived from a forensic sample. The Mitotyper Rules does not have these limitations because it is designed for forensic applications. The Mitotyper

TABLE 2—*Differences between types generated by EMPOP and by the Mitotyper Rules.*

Sample Name	EMPOP Alignment: rCRS and Type rCRS np 16180–16199	Mitotyper Alignment: rCRS and Type rCRS np 16180–16199
ITA0500260	AAAACCCCCTCC-CCATGCTT	AAAACCCCCTCCC-CATGCTT
SVN0600003	AAAACCCCCCCCCTCATGCTT	AAAACCCCCCCCCTCATGCTT
USA0600921	16189 T C	16189 T C
USA0601079	16191.1 - C 16192 C T	16192.1 - T
ESP0600264	AAAACCCCCTCCCCATGCTT	AAAACCCCCTCCCCATGCTT
HUN0600051	AAAACCTCCCCCC-ATGCTT 16186 C T 16189 T C 16193 C -	AAAACCTCC-CCCCATGCTT 16186 C T 16189 T -
USA0600888	AAAACCCCCTCCCC-ATGCTT	AAAACCCC-CTCCCCATGCTT
USA0600966	AAACCCCCTCCCCCCATGCTT 16183 A C 16188 C T 16189 T C 16193.1 - C	AAACCCCCTCCCCCCATGCTT 16183 A C 16187.1 - T 16189 T C
DNK0600006	AAAACCCCCTCCCC-A AACCCCCCTCCCCCCA 16182 A C 16183 A C 16188 C T 16189 T C 16193.1 - C	AAAACCCC-CTCCCCATGCTT AACCCCCCTCCCCCCATGCTT 16182 A C 16183 A C 16187.1 - T 16189 T C
ESP0600272	AAAACCCCCTCCCC-ATGCTT AAACCCCCCCCCTCCCATGCTT 16183 A C 16189 T C 16191 C T 16193.1 - C	AAAACCCCCTC-CCCATGCTT AAACCCCCCCCCTCCCATGCTT 16183 A C 16189 T C 16190.1 - T
ESP0600235	AAAACCCCCTCCCCATGCTT AAAACTCCCCCCC-ATGCTT 16185 C T 16189 T C 16193 C -	AAAACCCCCTCCCCATGCTT AAAACTCCC-CCCCATGCTT 16185 C T 16189 T -
POL0600375	AAAACCCCCTCCCC-ATGCTT AAAACCCCCCCCCTCATGCTT 16189 T C 16193 C T 16193.1 - C	AAAACCCCCTCCC-CATGCTT AAAACCCCCCCCCTCATGCTT 16189 T C 16192.1 - T
VNM0500147	AAAACCCCCTCCCCATGCTT ACCCCCCCCCCCC-ATGCTT 16181 A C 16182 A C 16183 A C 16189 T C 16193 C -	AAAACCCCCTCCCCATGCTT A-CCCCCCCCCCCCATGCTT 16181 A - 16182 A C 16183 A C 16189 T C
JPN0500063	AAAACCCCCTCCCC-ATGCTT AAACCCCCCCCCCCCCCCGCTT 16183 A C 16189 T C 16193.1 - C 16194 A C 16195 T C	AAAACCCCCTCCCCAT-GCTT AAACCCCCCCCCCCCCCCGCTT 16183 A C 16189 T C 16194 A C 16195 T C 16195.1 - C

Sample Name	np 37–51	np 37–51
POL0600267	AGCTCT-CCATGCATT AGCTCTCCCATGCATT 42.1 - C	AGCTCTCC-ATGCATT AGCTCTCCCATGCATT 44.1 - C

Sample Name	np 57–62	np 57–62
GRC0500023	TTTT-CG	T-TTTCG
GRC0500024	TCTTTCG	TCTTTCG
GRC0500047	58 T C	57.1 - C
HUN0500045	60.1 - T	
HUN0500135		
HUN0500193		
HUN0500250		
ITA0500300		
KEN0500066		

TABLE 2—*Continued*

Sample Name	EMPOP Alignment: rCRS and Type np 242–255	Mitotyper Alignment: rCRS and Type np 242–255
KEN0500065	CAATTGAATGTCTG	CAATTGAATGTCTG
KEN0500094	CAATTAA-TGTCTG	CAATT-AATGTCTG
USA0600925	247 G A 249 A -	247 G -

Sample Name	np 300–318	np 300–318
GRC0500158		
USA0600843	AAACCCCCCCTCCCCCGCT	AAACCCCCCCTCCCCCGCT
POL0600388	AAACCCCCCCCCCCC-GCT	AAACCCCCCC-CCCCCGCT
HUN0600362	310 T C 315 C -	310 T -
VNM0500039	AAACCCCCCCTCCCCCGCT	AAACCCCCCCTCCCCCGCT
VNM0500097	AAACCCCCCCCCCC--GCT 310 T C 314 C - 315 C -	AAACCCCCCC--CCCCCGCT 310 T - 311 C -
SVN0600092	AAACCCCCCC-TCCCCC-GCT AAACCCCCCTCTCCCCCCGCT 309 C T 309.1 - C 315.1 - C	AAACCCCCC-CTCCCCC-GCT AAACCCCCCTCTCCCCCCGCT 308.1 - T 315.1 - C
BEL0600071	AAACCCCCCC--TCCCCC-GCT AAACCCCTCCCCTCCCCCCGCT 307 C T 309.1 - C 309.2 - C 315.1 - C	AAACCCC-CCC-TCCCCC-GCT AAACCCCTCCCCTCCCCCCGCT 306.1 - T 309.1 - C 315.1 - C
GRC0500193	AAACCCCCCCTCCCCCGCT	AAACCCCCCCTCCCCCGCT
GRC0500219	AAACCCCCCCCCCCC--CT 310 T C 315 C - 316 G -	AAACCCCCCC-CCCCC-CT 310 T - 316 G -
HUN0500313	AAACCCCCCCTCCCCCGCT AAACCCCCCCCCCC--CT 310 T C 314 C - 315 C - 316 G -	AAACCCCCCCTCCCCCGCT AAACCCCCCC-CCCCC--T 310 T - 316 G - 317 C -

Sample Name	np 455–461	np 455–461
KEN0500015	T--CCCC-TC	T--CCCCTC
KEN0500056	TTTCCCCCTC 455.1 - T 455.2 - T 459.1 - C	TTTCCCCCTC 455.1 - T 455.2 - T 455.3 - C

Sample Name	np 513–526	np 513–526
HUN0600387	GCACACACACACCG	GCACACACACACCG
KEN0500047	ACACACACAC--CG	--ACACACACACCG
KEN0500056	513 G A	513 G -
KEN0500074	523 A -	514 C -
USA0600909	524 C -	
USA0600964		
GRC0500033	GCACACACACAC----CG GCGCACACACACACACCG 515 A G 524.1 - A 524.2 - C 524.3 - A 524.4 - C	GC----ACACACACACCG GCGCACACACACACACCG 514.1 - G 514.2 - C 514.3 - A 514.4 - C

Sample Name	np 568–582	np 568–582
DEU0600108	CCCCCC--ACAGTTTAT	CCCCCCAC--AGTTTAT
GRC0500242	CCCCCCCCCCCCAGTTTAT	CCCCCCCCCCCCAGTTTAT
GRC0500260	573.1 - C	574 A C
HUN0600168	573.2 - C 573.3 - C 574 A C	575.1 - C 575.2 - C 575.3 - C

TABLE 2—*Continued*

Sample Name	np 568–582	np 568–582
GRC0500316	CCCCCC--ACAGTTTAT CCCCCCCCCCAGTTTAT 573.1 - C 573.2 - C 574 A C	CCCCCCAC--AGTTTAT CCCCCCCCCCAGTTTAT 574 A C 575.1 - C 575.2 - C

EMPOP, European Mitochondrial DNA Population database; rCRS, revised Cambridge Reference Sequence.

TABLE 3—*Sequences found in EMPOP with multiple types.*

	EMPOP Alignment with rCRS and Type		EMPOP Alignment with rCRS and Type	Mitotyper Alignment with rCRS and Type
Sample Name	rCRS np 16180–16199	Sample Name	rCRS np 16180–16199	rCRS np 16180–16199
	CTCC-CCATGCTTACAAGCAA (rCRS)		CTCCCC-ATGCTTACAAGCAA (rCRS)	AAAACCCCCTCCC-CATGCTT (rCRS)
USA0600921	CCCCCTCATGCTTACAAGCAA	POL0600375	CCCCCTCATGCTTACAAGCAA	AAAACCCCCCCCCTCATGCTT
SVN0600003	16189 T C		16189 T C	16189 T C
ITA0500260	16191.1 - C		16193 C T	16192.1 - T
USA0601079	16192 C T		16193.1 - C	

Sample Name	rCRS np 300–318	Sample Name	rCRS np 300–318	rCRS np 300–318
	CAAACCCCCCC-TCCCCC-GCT (rCRS)		CAAACCCCCC-CTCCCCC-GCT (rCRS)	CAAACCCCCC-CTCCCCC-GCT (rCRS)
SVN0600092	CAAACCCCCCTCTCCCCCCGCT	HUN0500202	CAAACCCCCCTCTCCCCCCGCT	CAAACCCCCCTCTCCCCCCGCT
	309 C T		308.1 - T	308.1 - T
	309.1 - C		315.1 - C	315.1 - C
	315.1 - C			

Sample Name	rCRS np 37–51	Sample Name	rCRS np 37–51	rCRS np 37–51
	CTCACGGGAGCTCT-CCATGC (rCRS)		CTCACGGGAGCTCTCC-ATGC (rCRS)	CTCACGGGAGCTCTCC-ATGC (rCRS)
POL0600267	CTCACGGGAGCTCTCCCATGC	HUN0600207	CTCACGGGAGCTCTCCCATGC	CTCACGGGAGCTCTCCCATGC
	42.1 - C	AUT0500255	44.1 - C	44.1 - C
		KEN0500083		
		KEN0500041		
		KEN0500073		
		USA0600739		

Sample Name	rCRS np 57–62	Sample Name	rCRS np 57–62	rCRS np 57–62
	T-TTTC (rCRS)		TTTT-C (rCRS)	T-TTTC (rCRS)
GRC0500319	TCTTTC	HUN0500135	TCTTTC	TCTTTC
GRC0500123	57.1 - C	GRC0500023	58 T C	57.1 - C
		GRC0500024	60.1 - T	
		HUN0500250		
		HUN0500193		
		ITA0500300		
		GRC0500047		
		KEN0500066		
		HUN0500045		

EMPOP, European Mitochondrial DNA Population database; rCRS, revised Cambridge Reference Sequence.

Rules and the software implementing them provide consistency of nomenclature for the practitioner and provide a common functional system within and among laboratories. Therefore, we strongly recommend that all databases be reviewed with an automated system such as the Mitotyper software as a quality control to minimize inconsistencies.

Lastly, Wilson et al. (3) recommended full-text searching of an mtDNA type within a reference database to eliminate the effect of nomenclature/alignment inconsistencies (2). We continue to endorse the concept of full text for searching a database for matching sequences as it does not require any nomenclature or alignment strategy. However, full-text searching may not account for alignments that are one or more bases away from the searched sequence. Such alignment information may be useful to the forensic community when considering potential matches in missing person cases. In such cases, these searches are unlikely to occur through an EMPOP database search. Instead, the search will be in a database constructed of mtDNA sequences from personal items and maternal family members of the missing person. Alternate hypotheses to consider could be that the two sequences originate from two maternally related individuals and the difference is attributable to a mutation or that these two sequences arise from two unrelated

individuals. In this scenario, one could argue that an evolutionary approach would be a better model. However, if the sequence information is reduced, as might occur for degraded and contaminated samples, it may not be possible to derive a phylogenetic signature from the limited information. Therefore, consistency is still a guiding principle for forensic evidence. Regardless, the study herein demonstrates that both approaches, hierarchical and phylogenetic, are highly concordant and that results would likely be the same for the majority of haplotypes, even for kinship analyses.

Acknowledgments

This is publication number 08–16 of the Laboratory Division of the Federal Bureau of Investigation. Names of commercial manufacturers are provided for identification only, and inclusion does not imply endorsement by the Federal Bureau of Investigation. The authors thank Patricia Aagaard, Dr. Alice Isenberg, and Dr. Leslie McCurdy for their review and editorial contributions.

References

1. Andrews S, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet 1999;23:147.
2. Budowle B, Wilson MR, DiZinno JA, Fasano C, Holland MM, Monson KL. Mitochondrial DNA regions HVI and HVII population data. Forensic Sci Int 1999;103:23–35.
3. Wilson MR, Allard MW, Monson KL, Miller WP, Budowle B. Recommendations for consistent treatment of length variants in the human mtDNA control region. Forensic Sci Int 2002;129:35–42.
4. Scientific Working Group on DNA Analysis. Guidelines for mitochondrial DNA (mtDNA) nucleotide sequence interpretation (April 2003), Forensic Science Communication, http://www.fbi.gov/hq/lab/fsc/backissu/april2003/swgdammitodna.htm.
5. Bandelt HJ, Parson W. Consistent treatment of length variants in the human mtDNA control region: a reappraisal. Int J Legal Med 2008;122:11–21.
6. Budowle B, Fisher CL, Polanskey D, Den Hartog BK, Kepler RB, Elling JW. Stabilizing mtDNA sequence nomenclature with an operationally efficient approach. Forensic Sci Int Genet Suppl Ser 2008;1:671–3.
7. Budowle B, Polanskey D, Fisher CL, Den Hartog BK, Kepler RB, Elling JW. Automated alignment and nomenclature for consistent treatment of polymorphisms in the human mitochondrial DNA control region. J Forensic Sci. Online publication ahead of print. doi: 10.1111/j.1556-4029.2010.01478.x.
8. Mitochondrial DNA Control Region Database. Institute of Legal Medicine, Innsbruck Medical University, Innsbruck, Austria, http://www.empop.org (accessed April 25, 2008).

Additional information and reprint requests:
Deborah Polanskey, B.S.
FBI Laboratory
2501 Investigation Parkway
Quantico, VA 22135
E-mail: deborah.polanskey@ic.fbi.gov

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01478.x
Available online at: interscience.wiley.com

PAPER

CRIMINALISTICS

Bruce Budowle,[1] *Ph.D.; Deborah Polanskey,*[2] *B.S.; Constance L. Fisher,*[2] *Ph.D.; Bobi K. Den Hartog,*[3] *Ph.D.; Russell B. Kepler*[3]*; and John W. Elling,*[3] *Ph.D.*

Automated Alignment and Nomenclature for Consistent Treatment of Polymorphisms in the Human Mitochondrial DNA Control Region

ABSTRACT: Naming mtDNA sequences by listing only those sites that differ from a reference sequence is the standard practice for describing the observed variations. Consistency in nomenclature is desirable so that all sequences in a database that are concordant with an evidentiary sequence will be found for estimating the rarity of that profile. The operational alignment and nomenclature rules, i.e., "Wilson Rules," suggested for this purpose do not always guarantee a single consistent sequence description for all observed polymorphisms. In this work, the operational alignment/nomenclature rules were reconfigured to better reflect traditional user preferences. The rules for selecting alignments are described. In addition, to avoid human error and to more efficiently name mtDNA sequence variants, a computer-facilitated method of aligning mtDNA sample sequences with a reference sequence was developed. There were 33 differences between these hierarchical rules and the data in SWGDAM, which translates into a 99.92% consistency between the new rules and the manual historical nomenclature approach. The data support the reliability of the current SWGDAM database. As the few discrepancies were changed in favor of the new hierarchical rules, the quality of the SWGDAM database is further improved.

KEYWORDS: forensic science, mtDNA, nomenclature, hierarchal rules, parsimony, software, Mitotyper Rules™, phylogenetics, validation

Mitochondrial DNA (mtDNA) sequencing can be used effectively to analyze highly degraded or limited quantity forensic samples, particularly when nuclear DNA is not typeable (1). The sequencing of the noncoding control region (or only the hypervariable regions, HV1 and HV2) results in a string of approximately eleven hundred (or six hundred bases) or less if the sample is severely compromised. Displaying and conveying so many bases is cumbersome and impractical. Instead only those limited number of bases that differ from a published reference sequence (the revised Cambridge Reference Sequence, or rCRS (2,3) are described. Naming mtDNA sequences by listing only those sites that differ from the rCRS provides a common language and a simple operational tool for describing the variation observed. Identical or concordant sequences are identified by exhibiting the same differences from the rCRS.

This nomenclature concept of comparison to the rCRS has been well accepted but has a complication that needs to be addressed. This complication is the possibility of multiple rule-compliant alignments occurring against the rCRS for some sequences. A practical alignment scheme for naming the variants contained within the mtDNA sequences needs to be defined so practitioners can consistently select the same alignment option for a particular sequence. Such practice will provide nomenclature stability within and among forensic laboratories (4,5). Consistency of nomenclature is desired so that all sequences in a database that are concordant with an evidentiary sequence will be found for estimating the rarity of that profile. Wilson et al. (6,7) proposed an alignment and nomenclature protocol (the "Wilson Rules") to attempt to standardize the description of differences from the reference sequence. While the concept of the rules proffered by Wilson et al. was generally accepted, the rules themselves were not strictly practiced for several reasons. First, the rule favoring indels (i.e., insertions and deletions) over substitutions did not reflect some of the historically preferred alignments. [Note: we use historical to mean an alignment that was found in databases prior to implementation of any nomenclature rules]. Second, the nomenclature process was performed manually, which led to situations where the same sequence was named differently by different practitioners (8–10). Third, the rules as described did not always guarantee a single consistent sequence description at all observed polymorphic regions within an mtDNA sequence. Therefore, the alignment/nomenclature rules were reconfigured, and a computer-facilitated method of aligning mtDNA sample sequences with a reference sequence was developed. A software package named Mitotyper™ was created to implement the nomenclature rules that allows for rapid sequence alignment, provides for absolute stability and consistency, and provides increased database accuracy.

In this article, a hierarchical set of rules called the Mitotyper Rules™ are described. They were developed for greater consistency with both historical and current operational nomenclature. Additional rules were added to stabilize the operational nomenclature and to select among those few situations where two or more alignments are still possible after the modified operational rules have been executed. With this approach, absolute consistency and

[1]University of North Texas, HSC, 3500 Camp Bowie Blvd., Fort Worth, TX.
[2]FBI Laboratory, 2501 Investigation Parkway, Quantico, VA.
[3]Mitotech LLC, 590 Monte Alto, Santa Fe, NM.

Received 8 Jan. 2009; and in revised form 12 May 2009; accepted 31 July 2009.

stability in nomenclature is attained. Those haplotypes, once named and entered into the database never need to be renamed; identical sequences will always be listed the same way, and database searching for statistical inferences will be greatly facilitated.

Software Operation and Alignment Strategy

The mtDNA type generation software first roughly aligns the reference rCRS and the sample sequence to identify large regions in which the sequences "match" one another. The Smith-Waterman-Gotoh algorithm is used for this rough alignment (11). Regions between matching regions are then, by our definition, polymorphic regions within which the sample sequence needs to be defined in relation to the rCRS. Polymorphic regions will contain one or a set of closely spaced sequence variants, the placement of which interact with one another according to the alignment and nomenclature rules.

The boundaries of the polymorphic regions are adjusted to include contextual information about neighboring bases for use when aligning and typing the region. This adjustment can include expansion or subdivision. Expansion can take place, for example, when maintaining a repeat motif in a contiguous fashion. Subdivision occurs, for example, around the HV2 C-stretch area to ensure alignment between specific landmarks of the polymorphic region.

In each polymorphic region, the set of every possible parsimonious alignment between the reference and the sample sequence is created. This set of possible alignments for each region is then processed by applying each rule in series to select the compliant alignments in the set. The standard execution orders of the rules are shown in Fig. 1, and the additional modified procedure used for aligning to the rCRS 300 to 315 region is shown in Fig. 2.

The software uses each rule to discard all of the alignments in the region's set of possible alignments that do not comply with the rule. If only one alignment remains after discarding the noncompliant alignments that alignment is used to create the type description of the polymorphism(s) in that region. However, after running the rule if more than one valid alignment still remains in the set of alignments for that region, the software uses the next rule in the series for processing the set.

Typing Rules and Procedures

The software applies the rules to analyze sets of candidate alignments in the hierarchy described in Figs. 1 and 2.

Rule 1—Least Number of Differences

The general overarching principle for the nomenclature rules is to select an alignment with the least number of differences between the rCRS and the sample sequence; that is, the most parsimonious alignment (6,7). For each polymorphic region, the software only generates the set of all possible alignments of the sample sequence in the region that have the minimum number of differences from the rCRS.

An ambiguous or heteroplasmically equivalent substitution in the sample consensus sequence is not considered a difference from the rCRS. The International Union of Pure and Applied Chemistry (IUPAC) nomenclature is used to determine heteroplasmic equivalence. For example, an A is considered equivalent to an R, W, M, D, H, V, or N but not a Y (12). An inserted ambiguous base in the sample sequence is counted as a difference from the rCRS.

Rule 2—Maintain the AC Repeat Motif

The rCRS HVIII region includes an AC repeat motif in the 515–525 range (13,14). If the polymorphic region includes the AC repeats, the software selects alignments that both place insertions or

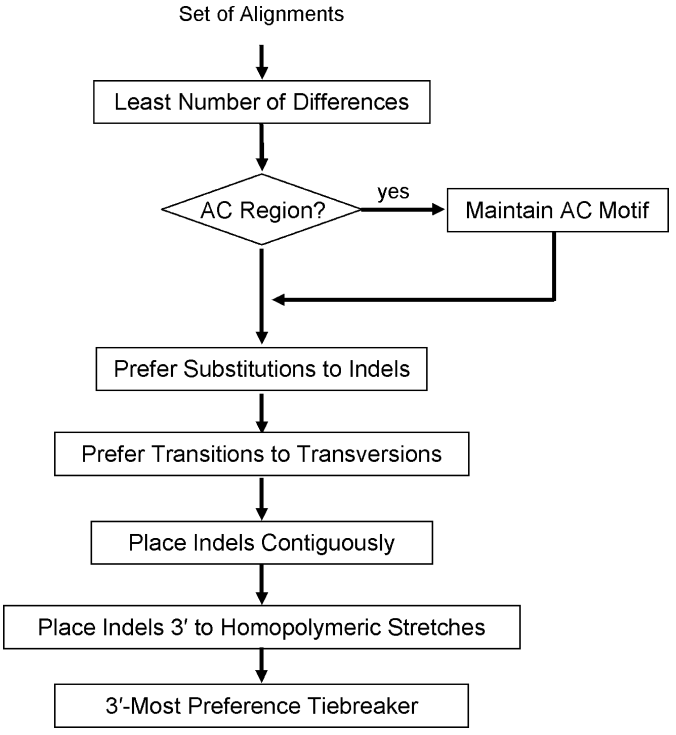


FIG. 1—*Standard application of the rules to the non-HV2CS regions containing poymorphisms.*

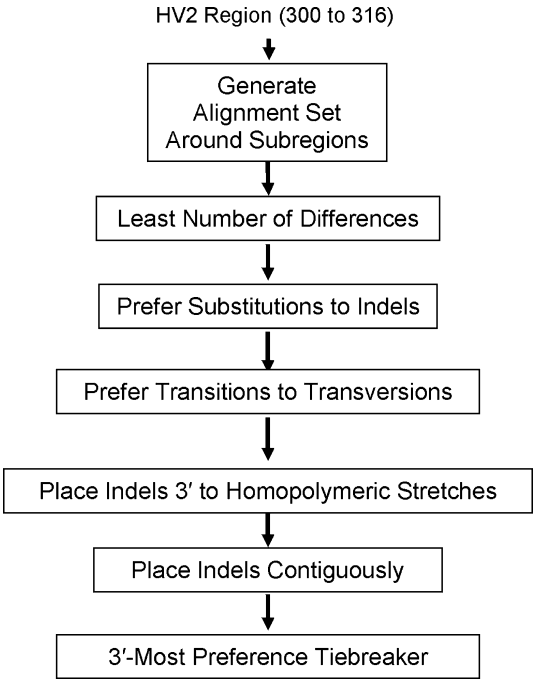


FIG. 2—*The HV2CS Typing Protocol, used for alignments in the 300–315 region, that anchors a T at position 310, if possible, and anchors the boundaries of subregions.*

deletions (indels) at the 3′end of the region and preserve the "AC" motif from those that place indels in less preferred positions. Table 1 contains an example of a historical alignment and a MitoTyper alignment in which the two deletions are placed together to reflect the deletion of one AC motif in the sample sequence, generating a preferred type of 523 A –, 524 C – in this region.

Rule 3—Prefer Substitutions to Indels

Polymorphisms between sequences can be aligned by describing the difference in bases as a substitution(s) or an indel(s). A substitution occurs when sequences are aligned so that one base is replaced with another to describe the polymorphism. Indels can be used to describe the same polymorphism, by deleting the mismatched base and inserting the sequence base that is observed, or some variation thereof.

This rule dictates a preference for alignments with substitutions rather than indels. The software chooses between candidate alignments by selecting the alignment with the most substitutions.

Table 2 shows an example of a preference for substitutions over indels to describe a polymorphism and contrasts that to a historical alignment found in the Scientific Working Group on DNA Analysis Methods (*SWGDAM*) database. This historical alignment was described by three indels while the Mitotyper alignment is achieved with two substitutions and one indel: 16183 A C 16187.1 – T, and 16189 T C.

Rule 4—Prefer Transitions to Transversions

Nucleotide substitutions can be classified as transitions or transversions. Generally, transitions are more biologically feasible because the substituted nucleotides belong to the same chemical group, purine or pyrimidine. In contrast, transversions involve nucleotide substitution between the purine and pyrimidine groups. This rule selects those alignments in which polymorphisms are described by transitions. Table 3 displays a historical alignment (that appears to favor transversions over transitions) in which the polymorphic region was described as an indel 56.1 – C, a transversion 57 T G, and a transition 73 A G. The software instead selected an alignment that described the polymorphisms with an indel and two transitions 57 T C, 57.1 – G, and 73 A G.

TABLE 1—*Maintain AC motif.*

	Historical	Mitotyper Rules (Preferred)
rCRS 512–527	AGCACACACACACCGC	AGCACACACACACCGC
Sample consensus	AGCACACACAC-C-GC	AGCACACACAC--CGC

TABLE 2—*Preference for substitutions to indels.*

	Historical	Mitotyper Rules (Preferred)
rCRS 16178–16193	TCAAAACCCCCTCCCC--	TCAAAACCCC-CTCCCC
Sample consensus	TCAAA-CCCCCTCCCCCC	TCAAACCCCCTCCCCCC

Rule 5—Place Indels Contiguously When Possible

Indels, at times, can be placed in several locations to align polymorphic sequences. Placement of indels at multiple locations is particularly possible in homopolymeric regions. The preferred alignment for nomenclature consistency places the indels in continuous groups as much as is possible. This rule is implemented by selecting between two candidate alignments the one with the fewest indel groups (i.e., groups of one or more contiguous indels). Table 4 shows an example of a historical type and a preferred type identified by the software where the same three indels can be placed contiguous to one another as 455.1 – T, 455.2 – T, and 455.3 – C.

Rule 6—Place Indels 3′ to Homopolymeric Stretches

The Wilson Rules describe basic alignment procedures for length variants such as placing indels and gaps at the 3′ end of the region when variation occurs in homopolymeric regions (6,7). This preference also is implemented in the Mitotyper software by selecting alignments that preserve homopolymeric stretches of base sequence and place the indels at the 3′ end of the homopolymeric stretch. Tables 5 and 6 illustrate examples of two historic alignments found in the SWGDAM data in which indels were not placed 3′ to homopolymeric stretches (i.e., incorrectly placed) along with the preferred alignments that comply with this rule. In the historical example in Table 5, the indel is at the wrong end of the poly-A stretch (i.e., the 5′ end). In the example in Table 6, the preferred alignment creates a two-T homopolymeric stretch.

Rule 7—Three Prime Preference Tiebreaker

It is possible for the set of alignments for each region to have been processed by all of the previous rules and still have multiple valid candidate alignments remaining. Rule 7 implements the general 3′ preference (15) as a tiebreaker to select the single preferred alignment for each region. For a set of alignments to reach this point, each alignment will have the same number of indel groups but not have both insertions and deletions in exactly the same positions in any two alignments.

The Mitotyper software first selects between two candidate alignments the one with the insertion or insertions at the 3′ end of the sequence. If a tie remains because two or more alignments have an insertion at the same 3′ most position, the software selects between these by preferring the alignment that has the 3′ most deletion.

Table 7 shows two alignments of a sample sequence that are both compliant with all previous rules. The insertion tiebreaker rule leads to the selection of the alignment with an insertion at 16193 over the alternative alignment that placed an insertion at position 16188.

TABLE 4—*Example of the place indels contiguously when possible rule.*

	Historical	Mitotyper Rules (Preferred)
rCRS 451–460	ATTTT--CCCC-T	ATTTT---CCCCT
Sample consensus	ATTTTTTCCCCCT	ATTTTTTCCCCCT

TABLE 3—*Preference for transitions to transversions.*

	Historical	Mitotyper Rules (Preferred)
rCRS 55–75	TA-TTTTCGTCTGGGGGGTATG	TAT-TTTCGTCTGGGGGGTATG
Sample consensus	TACGTTTCGTCTGGGGGGTGTG	TACGTTTCGTCTGGGGGGTGTG

TABLE 5—*Placing indels 3′ to a homopolymeric stretch, example 1.*

	Historical	Mitotyper Rules (Preferred)	Preferred Type
rCRS 16170–16183	AATCCACATCAAAA	AATCCACATCAAAA	
Sample consensus	AACCCACATC-AAA	AACCCACATCAAA-	16172 T C 16183 A –

TABLE 6—*Placing indels 3′ to a homopolymeric stretch, example 2.*

	Historical	Mitotyper Rules (Preferred)	Preferred Type
rCRS 111–114	CCCT	CCCT	
Sample consensus	CT-T	C-TT	113 C – 114 C T

TABLE 7—*Example of the tiebreaker rule using insertions.*

	Alignment 1	Alignment 2 (Preferred)
rCRS 16180–16194	AAAACCCCC-TCCCC	AAAACCCCCTCCCC-
Sample consensus	AAACCCCCCCCCCCC	AAACCCCCCCCCCCC

TABLE 8—*Example of the tiebreaker rule using deletions.*

	Historical	Mitotyper (Preferred)
rCRS 16180–16194	AAAACCCCCTCCCCA	AAAACCCCCTCCCCA
Sample consensus	AAAACCCC-ACCCCA	AAAACCCCA-CCCCA

Table 8 displays another example of two alignments that are both compliant with all previous rules. The Rule 7 Tiebreaker leads to the selection of the alignment with a deletion at 16189 over the historical alignment that placed the deletion at position 16188.

HV2 C-stretch (HV2CS) Region Typing

The five cytosine residues in the rCRS in the HV2 region from positions 311–315 represent a rare configuration from the majority of the population that has six cytosines in this C-stretch region (2). As a result, almost 99% of the sequences in the SWGDAM database (4834 of the 4839 sequences tested) have a 315.1 C type in this region.

Analysis of the SWGDAM and European mtDNA Population (EMPOP) databases reveals a preference for aligning the thymine at the 310 position (the 310T) in the rCRS with a thymine residue in the sample sequence and preserving the 315.1 C type description for polymorphisms in this region. Table 9 shows the sequence found in four SWGDAM samples (CHN.ASN.000091, CHN.ASN.000451, THA.ASN.000021, and USA.335.000132) and two EMPOP samples (GRC0500109 and GRC0500120) and their corresponding historical type in this region. This sequence is described with three indels, 308 C –, 309 C –, and 315.1 – C, which preserves the 310 T alignment. However, application of the Mitotyper Rules that follow the hierarchy shown in Fig. 1 would produce an alternate nonpreferred alignment, which would have two differences from the rCRS (“308 C T” and “310 T –”), thus meeting the “least number of differences” requirement.

Analysis of the historical data also reveals a preference for aligning the 300–302 poly-A region and the 303–309 poly-C region in the rCRS with matching regions in the sample sequence, also at the expense of a strict Rule 1 compliance. Table 10 shows the sequence from CHN.ASN.000451 with the preferred historical alignment (299 C –, 309.1 – C, 315.1 – C) type and the nonpreferred alignment produced by application of the Mitotyper Standard Rules in Fig. 1.

To correspond with historical preferences, typing polymorphisms in the HV2 300–315 region is carried out with a slightly different procedure shown in Fig. 2.

The HV2CS typing protocol begins by isolating this region from the rest of the sequence and generating a set of alignments between the sample and the reference sequence in the region that aligns the rCRS 310T with the candidate thymine(s) in the sample sequence (if it exists). It also aligns the 5′ 300 and 3′ 302 positions of the AAA subregion and the ending 3′ 315 boundary in the reference sequence with the corresponding candidate subregions in the sample sequence. In the case where there are ambiguous bases on the boundary, the software creates several alternatives that group the ambiguous bases in one subregion or the other.

With this procedure, each alignment in the set for this region will reflect the traditional practice of aligning T residues and the boundaries of the poly-A and poly-C subregions. This set of alignments is then processed by sequential application of the typing rules applied in the order illustrated in Fig. 2. As before, each rule is responsible for discarding all of the alignments in the regions' set of possible alignments that do not comply with the rule.

When typing the HV2CS, the order in which the rules “place indels contiguously” and “place indels 3′ to homopolymeric

TABLE 9—*Historical preference for aligning the 310T in the rCRS 300–315 region.*

	Historical and Mitotyper HV2CS Results (Preferred)	Mitotyper Standard Rules
rCRS 295–317	CCACCAAACCCCCCCTCCCCC-GC	CCACCAAACCCCCCCTCCCCCG
Sample consensus	CCACCAAACCCCC--TCCCCCCGC	CCACCAAACCCCCTC-CCCCCGC

TABLE 10—*Historical preference for aligning the poly-A and poly-C regions in the rCRS.*

	Historical and Mitotyper HV2CS Results (Preferred)	Mitotyper Standard Rules
rCRS 295–317	CCACCAAACCCCCCC-TCCCCC-GC	CCACCAAACCCCCCCTCCCCC-GC
Sample consensus	CCAC-AAACCCCCCCCTCCCCCCGC	CCACAAACCCCCCCCTCCCCCCGC

TABLE 11—*Historical preference for aligning the poly-A and poly-C regions in the rCRS.*

	Mitotyper HV2CS Rules (Preferred)	Mitotyper Standard Rules
rCRS 300–317	AAACCCCCCCCT-CCCCC-GC	AAACCCCCCCT–CCCCCGC
Sample consensus	AAACCCCCCCTTCCCCCCGC	AAACCCCCCCTTCCCCCCGC

stretches'' are applied is reversed from the standard Mitotyper rules (Fig. 1). This change preserves the preferred placement of the sixth C after the 310T as 315.1C in the event that there is a second insertion in the sample in the 310–315 subregion of the HV2CS. Table 11 illustrates an example of the effect of this change. In the standard prioritization, ''Place Indels Contiguously When Possible,'' first discards any alignments in which the desired 315.1C indel is separate from another indel in this subregion. Instead, prioritizing the rule ''Place Indels 3′ to Homopolymeric Stretches'' first in the HV2CS typing procedure results in the selection of the historically preferred alignment of the 315.1C insertion at the end of the poly-C region with the additional 310.1—T type.

Results

The Mitotyper Rules were derived from the principle of a hierarchical operational alignment approach proposed by Wilson et al. (6,7). This development process resulted in nomenclature rules that are familiar to the forensic analyst and maximally congruent with historic practices. The Mitotyper Software™ was produced to automate the alignment and nomenclature described by the rules herein for selecting consistent alignments of mtDNA sequences with the rCRS and generating type descriptions of polymorphisms for forensic and statistical use.

Testing and Validation

The SWGDAM mtDNA population database was used as the historical reference data set when developing the Mitotyper Rules (16). In the development process, these historical types were compared with the Mitotyper rule-compliant type generated for each sample.

In the 4839 SWGDAM sequences, the software identified 40,357 polymorphic regions that contained one or more differences from the rCRS. After typing with the software, only 33 of the 40,357 regions in 33 samples received types that did not agree with the historical types—a notably high 99.92% consistency considering that a manual nomenclature typing procedure was used to generate the database.

Of the 40,357 polymorphic regions, 32,915 had only one alignment between the sample and the rCRS with the minimum number of differences and so the type was determined solely by the Rule 1 parsimony requirement. For the remaining 7442 polymorphic regions (distributed across 4802 samples), the software generated a set of candidate alignments all with the same minimum number of differences, which needed to be processed by subsequent rules.

An additional 1254 samples served as an independent validation data set. The typing software found 11,303 polymorphisms in this set of haplotypes; only 20 different sequences (with 56 occurrences) were discordant with the Mitotyper software types. In 11 of the 20 cases, the software produced a more parsimonious type. Manual examination of the 56 discrepancies by qualified analysts resolved these differences in favor of the software type. The low 0.18% discrepancy rate, similar to the rate achieved in the SWGDAM data set, indicates that the Mitotyper rules were not over fit to the SWGDAM data.

Mitochondrial DNA sequences of HV1 and HV2 of 1204 unrelated Japanese individuals were used as an additional independent validation data set (17). The Mitotyper software found 8859 polymorphic regions in the 741 different haplotypes in the data set. Five polymorphic subsequences that occurred in 11 samples were typed differently between the historical type and the Mitotyper result. The 99.88% congruence illustrates the utility of the Mitotyper rules and automated software to support the currently practiced nomenclature for describing mtDNA sequence polymorphisms.

Software Testing—Robustness

The robustness and reliability of the Mitotyper software was tested on over six million computer-generated sample sequences. Each test sample sequence was generated by introducing a random number of variants (between 1 and 30) to the reference sequence control region from nucleotide positions 16025 to 525. The modification location was random with an equal distribution across the sequence, and the modification was randomly chosen with an 80% chance of being a substitution and 10% chance each of being an insertion or a deletion. An additional heteroplasmic or ambiguous modification was inserted at a random location by replacing the base with an IUPAC symbol in the set ''K'',''M'',''R'',''D'',''S'',''H'',''W'',''V'',''Y'',''B'', or ''N''. In every case, the type generation software returned a type description that accurately described the computer-generated polymorphisms that had been randomly generated.

Conclusions

The Mitotyper Rules described here are familiar to and easily followed by the forensic analyst. These rules are stable, well defined, and maximally consistent with both historical and current operational nomenclature and can be applied manually. The Mitotyper software encodes the Mitotyper Rules, automating the description of sequence differences in relation to the rCRS. While the software is not strictly required to follow the rules described herein, implementing these hierarchical rules using robust software removes human involvement in selecting alignments, enhances consistency, and offers substantial savings in time and effort to describe mtDNA sequence polymorphisms.

The 33 differences between the software results and the historical types that were observed in the 4839 samples in the public SWGDAM database are resolved in favor of the software type. This updated version of the SWGDAM database will be available at http://www.fbi.gov/hq/lab/fsc/backissu/april2003/swgdammitodna.htm. All future sample types in SWGDAM, including the additional 1254 samples used in the validation study, will be compliant with the Mitotyper rules.

It is important for the forensic community to adopt a strategy that resolves ambiguities and selects consistently a particular alignment for employing mtDNA nomenclature. These recommendations herein stabilize nomenclature, provide for higher quality databases, and enable better, more reliable profile search results for evaluating the weight of evidence. Conceivably, as more samples are typed, a novel polymorphism may require additional rules; these can be

added to the software tool, but previously typed mtDNA profiles will not have to be renamed (15).

Acknowledgments

This is publication number 08–17 of the Laboratory Division of the Federal Bureau of Investigation. Names of commercial manufacturers are provided for identification only, and inclusion does not imply endorsement by the Federal Bureau of Investigation.

We thank K. Sekiguchi, K. Imaizumi, K. Fujii, N. Mizuno, Y. Ogawa, T. Akutsu, H. Nakahara, T. Kitayama, and K. Kasai of National Research Institute of Police Science, Kashiwa-shi, Chiba, Japan, for kindly providing their population sequence data for our studies. We thank Patricia Aagaard, Dr. Susan Cropp and Dr. Alice Isenberg for their editorial contributions.

References

1. Wilson MR, DiZinno JA, Polanskey D, Replogle J, Budowle B. Validation of mitochondrial DNA sequencing for forensic casework analysis. Int J Legal Med 1999;108:68–74.
2. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet 1999;23:147.
3. Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, et al. Sequence and organization of the human mitochondrial genome. Nature 1981;290:457–65.
4. Bender K, Schneider PM, Rittner C. Application of MtDNA sequence analysis in forensic casework for the identification of human remains. Forensic Sci Int 2000;113:103–7.
5. Szibor R, Michael M, Plate I, Krause D. Efficiency of forensic mtDNA analysis. Case examples demonstrating the identification of traces. Forensic Sci Int 2000;113:71–8.
6. Wilson MR, Allard MW, Monson KL, Miller KWP, Budowle B. Recommendations for consistent treatment of length variants in the human mitochondrial DNA control region. Forensic Sci Int 2002;129:35–42.
7. Wilson MR, Allard MW, Monson KL, Miller KWP, Budowle B. Further discussion of the consistent treatment of length variants in the human mitochondrial DNA control region. Forensic Science Communication, October 2002;4(4), http:// ww.fbi.gov/hq/lab/fsc/current/index.htm.
8. Carracedo A, D'Aloja E, Dupuy B, Jangblad A, Karjalainen M, Lambert C, et al. Reproducibility of mtDNA analysis between laboratories: a report of the European DNA Profiling Group (EDNAP). Forensic Sci Int 1998;97:165–70.
9. Yao YG, Bravi CM, Bandelt H-J. A call for mtDNA data quality control in forensic science. Forensic Sci Int 2004;141:1–6.
10. Budowle B, Polanskey D, Allard MW, Chakraborty R. Addressing the use of phylogenetics for identification of sequences in error in the SWGDAM mitochondrial DNA database. J Forensic Sci 2004;49:1256–61.
11. Moustafa A, Aligner J. Open source Java implementation of Smith-Waterman, http://jaligner.sourceforge.net.
12. http://www.bioinformatics.org/sms/iupac.html.
13. Bodenteich A, Mitchell LG, Polymeropoulos MH, Merril CR. Dinucleotide repeat in the human mitochondrial d-loop. Hum Mol Genet 1992;1:140.
14. Bodenteich A, Mitchell LG, Merril CR. A lifetime of retinal light exposure does not appear to increase mitochondrial mutations. Gene 1991;108:305–10.
15. Budowle B, Fisher CL, Polanskey D, Den Hartog BK, Kepler RB, Elling JW. Stabilizing mtDNA sequence nomenclature with an operationally efficient approach. Forensic Sci Int Genetics Supplement Series 2008;1:671–3.
16. Monson KL, Miller KWP, Wilson MR, DiZinno JA, Budowle B. The mtDNA population database: an integrated software and database resource for forensic comparison. Forensic Science Communication 2002;4(2), http://www.fbi.gov/hq/lab/fsc/backissu/april2002/miller1.htm.
17. Sekiguchi K, Imaizumi K, Fujii K, Mizuno N, Ogawa Y, Akutsu T, et al. Mitochondrial DNA population data of HV1 and HV2 sequences from Japanese individuals. Leg Med 2008;10:284–6.

Additional information and reprint requests:
Deborah Polanskey, B.S.,
FBI Laboratory, 2501 Investigation Parkway
Quantico, VA
E-mail: deborah.polanskey@ic.fbi.gov

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01440.x
Available online at: interscience.wiley.com

PAPER

CRIMINALISTICS

Sergio Cardoso,[1] *Ph.D.; María T. Zarrabeitia,*[2] *Ph.D.; Laura Valverde,*[1] *B.Sc.; Adrian Odriozola,*[1] *B.Sc.; Miguel Á. Alfonso-Sánchez,*[2] *Ph.D.; and Marian M. de Pancorbo,*[1] *Ph.D.*

Variability of the Entire Mitochondrial DNA Control Region in a Human Isolate from the Pas Valley (Northern Spain)

ABSTRACT: In this study, we analyzed the entire mtDNA control region in 61 unrelated individuals from the Pas Valley (Cantabria), a human isolate from northern Spain, to evaluate the suitability of this analysis to increase the power of discrimination of this locus for forensic purposes in human isolates. Low values obtained for the diversity parameters confirmed the relative isolation of this human group. The main findings of this study indicated that even the analysis of the entire mtDNA control region may have important limitations for use in forensic casework when dealing with human isolates: none of the 44 individuals who exhibited identical HVI-HVII haplotypes could be further differentiated by analysis of segment HVIII. Nevertheless, analysis of the entire mtDNA control region proved to be useful to determine the ancestry of the samples examined, by contributing to the confirmation, and, on occasion, even to the refinement of the haplogroup assignment.

KEYWORDS: forensic science, mitochondrial DNA, D-loop region, human isolate, haplogroup, Pas Valley

Hypervariable segments HVI and HVII of the mitochondrial DNA (mtDNA) control region have been largely used in forensic casework analysis (1), even when the power of discrimination provided by mtDNA sequencing is clearly below the values obtained from autosomal short tandem repeats (STRs). However, when the amount of nuclear DNA is restraining or when the sample is highly degraded, mtDNA arises as the only option to obtain genetic data (2). In an attempt to overcome this limitation, Lutz et al. (3) proposed that analysis of segment HVIII of the mtDNA might be useful to differentiate between individuals harboring the same HVI-HVII haplotype. This issue was further investigated in subsequent studies that confirmed the potential effectiveness of hypervariable segment HVIII (4), and also of the entire mtDNA control region, as an approach to increase the resolving power of the mtDNA for forensic applications (5).

The extremely scarce information available regarding HVIII or the entire mtDNA control region includes large panmictic populations such as Germany (3), Bologna (4), or Japan (5). However, the analysis of the entire mtDNA control region remains uncovered in isolated human groups. Human isolates stand out by low sequence diversity values, and, therefore, analysis of HVI and HVII sequences renders a lower power of discrimination relative to panmictic populations. This is the case of the Basque subpopulations from the Basque Country (6) and northern Navarre (7), two relatively isolated human groups settled in the Franco-Cantabrian area that feature low sequence diversity (considering HVI-HVII haplotypes) with respect to other European populations.

In this study we present a detailed analysis of the entire mtDNA control region in a population sample from the Pas Valley, a human isolate from Cantabria (northern Spain). The autonomous community of Cantabria comprises a population of about 560,000 inhabitants over an area of *c.* 5300 km^2, most of them living in the northern coastal fringe. The Pas Valley is located in the south part of the community and is characterized by a landscape of valleys, separated from each other by forests and mountains. The economic mainstay of the Pas Valley is cattle breeding, mostly conditioned by the abrupt orography of this area. Thus, autochthonous people have a very ancient tradition as shepherds, characterized by the practice of the altitudinal transhumance: the shepherds move over short distances, carrying their livestock to pastures in high or low valleys depending on the season. The genetic isolation of the Pas Valley region has been confirmed by high endogamy and consanguinity levels (8). From the genetic viewpoint, analysis of Y-chromosome haplotypes and mtDNA HVI sequences (9,10) revealed a high degree of genetic differentiation of Pasiegos with respect to other Iberian populations. Thus, this population represents an exceptional human group to corroborate the potential usefulness of analyzing HVIII together with HVI and HVII, or even the entire control region of mtDNA in the field of forensics.

The main objective of this study was to determine the suitability of analyzing the entire mtDNA control region in human isolates to increase the power of discrimination of this locus in forensic casework. On the other hand, we also aimed at contributing new mtDNA haplotype data, because the improvement of databases constitutes a major goal to consolidate the use of mtDNA for forensic purposes.

[1]BIOMICS Research Group, Centro de Investigación y Estudios Avanzados "Lucio Lascaray," Universidad del País Vasco, Miguel de Unamuno 3, 01006, Vitoria-Gasteiz, Spain.

[2]Unidad de Medicina Legal, Universidad de Cantabria, Av Herrera Oria s/n, 39011 Santander, Spain.

Received 8 May 2009 and accepted 4 July 2009.

Materials and Methods

Sample Collection

Blood samples were collected from 61 healthy and maternally unrelated individuals living in two of the main demographic centers of the Pas Valley (Cantabria province, northern Spain): Vega de Pas ($N = 31$) and San Pedro del Romeral ($N = 30$). Three generation pedigree charts were compiled to guarantee the collection from autochthonous, nonrelated individuals. Thus, all participants were asked about the geographic origins of their parents and grandparents to confirm maternal origin in the Pas Valley. All donors gave their informed consent prior to inclusion in the sample, following the ethical guidelines stipulated by each of the institutions involved in the study.

PCR Amplification and Sequencing

Genomic DNA was extracted from peripheral blood by means of QIAamp DNA blood mini kit (Qiagen, Hilden, Germany), following manufacturer guidelines. The entire control region of mitochondrial DNA was PCR amplified using the primer set L15988 (15965-15987, 5′-AAGTCTTTAACTCCACCATTAGC-3′) and H616 (637-617, 5′-GTGATGTGAGCCCGTCTAAAC-3′). PCRs were performed in an iCycler (Bio-Rad Lab., Hercules, CA). The reactions were carried out in a final volume of 25 μL containing 0.05 mM dNTP Master Mix (Bioline), 1.5 mM $MgCl_2$ (Bioline), 10 × buffer (Bioline), 0.1 μM of each primer, 0.01 mg of BSA (Roche), and 0.75 U of BioTaq polymerase (Bioline). The DNA template was 10 ng of dsDNA quantified by means of Quant-iT™ PicoGreen® dsDNA Assay Kit (Invitrogen, Carlsbad, CA). The PCR consisted of 3-min denaturation at 95°C, 10 cycles of 45-s denaturation at 95°C, 1-min annealing at 70°C decreasing 1°C/cycle and 90-s extension at 72°C, 25 cycles of 45-s denaturation at 95°C, 1-min annealing at 60°C, a 90-s extension at 72°C, and a final 10-min extension at 72°C. PCR products were electrophoresed on 1.5% agarose gels to check for correct amplification. These products were purified by vacuum ultrafiltration using MultiScreen® PCRμ96 plates (Millipore, Billerica, MA) and sequenced using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA) on an ABI3130 Genetic Analyzer (Applied Biosystems). We performed sequencing using primers L15988 and L12 (5′-ACATCACGATGGATCACAGGTC-3′). Samples with poor resolution downstream of C-stretches were resequenced using primer H285 (5′-GGGGTTTGGTGGAAATTTTTTG-3′) and/or H616.

Statistical Analysis and Haplogroup Classification

MtDNA sequences were edited using Sequence Scanner software v1.0 (Applied Biosystems). These sequences were subsequently aligned and compared with the revised Cambridge reference sequence (rCRS) (11) using Clustal X v1.83 program package (12). Sequences were edited between np 16024 and np 576. All the polymorphic positions detected were confirmed in chromatograms. Within-population variation in mtDNA sequences was analyzed by calculating diversity parameters in the Arlequin v3.1 program (13): nucleotide diversity (π) (14,15), gene diversity (H) (15), and the mean number of pairwise differences ($\theta\pi$) (14). The probability of randomly selecting two samples with the same mitochondrial profile (random match probability) was calculated using the equation $p = \sum X^2$, where x is the frequency of each mtDNA haplotype (16). Because length variations of the mtDNA sequences are misleading for interpretation purposes (17), C-stretch variation and heteroplasmy were excluded from statistical analyses. Finally, haplotypes were classified into haplogroups as described in Alfonso-Sánchez et al. (6).

Results and Discussion

Polymorphisms in the mtDNA Control Region

The complete mtDNA control region, including hypervariable segments HVI, HVII, and HVIII, was sequenced in 61 unrelated individuals from the Pas Valley (Spain). The resultant mitochondrial DNA sequences have been deposited at the GenBank database under accession numbers FJ800302-FJ800362. The sequences were aligned and compared with the rCRS (11). DNA sequencing revealed 66 polymorphic sites for the entire control region (Table 1), 30 of them located in HVI (45.5%), 21 in HVII (31.8%), and 10 in HVIII (15.2%).

All the polymorphisms observed in HVI and HVII were substitutions (Table 2). We also observed insertions in the poly-C stretches located in both segments, at positions 16193 and 309, respectively. The polymorphic positions observed in HVIII were similar to those described by Lutz et al. (18) and Bini et al. (4) for Germany and Bologna, respectively. Additionally, segments located between hypervariable sequences showed five polymorphic sites in our collection. In all, polymorphic sites identified in HVIII and non-HV sequences represented the 22.7% of the variable sites within the entire control region. This result is in good agreement with the variability reported for a Japanese sample (5).

Indels represented only a 7.6% of total polymorphisms observed in our study, whereas 92.4% were substitutions. In addition, we found 8 insertions in poly-C stretches. Segment HVI showed the highest proportion of substitutions (49.2% over total substitutions), whereas indels appeared solely in HVIII. These indels were insertions or deletions of CA-repeats between nps 514–523. We found 47 sequences with $(CA)_5$, 11 with $(CA)_{5\text{-}1}$, two sequences with $(CA)_1(CT)_1(CA)_4$, and finally one sequence with $(CA)_{5+1}$ repeats. Likewise, we found an insertion of a thymine at np 455 in one individual, a phenomenon not reported in previous studies focused on HVIII (4,5,18). These studies found transition C456T at moderate frequencies (Germany: 4.5%, Bologna: 3.0%, Japan: 1.6%); however, this nucleotide substitution was not observed in the Pas Valley sample. Interestingly, the insertion at np 455 was observed in two complete genomes reported by Palanichamy et al. (19). Furthermore, these two individuals and our sample shared polymorphisms at np 16129, 16223, 16311, 16391, 16519, 73, 199, 204, 250, and 573 in the mtDNA control region.

Similarly to HVI and HVII, segment HVIII possesses a homopolymeric C-stretch between nps 568–573. As noticed in previous publications (4,18), this nucleotide structure promotes length heteroplasmy in HVIII, albeit at a lower frequency than in segments HVI and HVII. This heteroplasmy was observed in only one of the individuals analyzed herein.

Haplotypes and Forensic Estimates

Analysis of HVI sequences obtained for the 61 individuals from the Pas Valley permitted the identification of 22 different haplotypes defined by 30 segregating nucleotide positions (see Table 1). Of the whole set of HVI haplotypes, 8 sequences were unique (13.1%) and 5 sequences (8.2%) were identical to the rCRS. On the other hand, a total of 21 HVII haplotypes were observed, defined by 21 segregating nucleotide positions. Of them, 13 sequences were unique (21.3%) and, unlike HVI, no HVII

TABLE 1—*Haplogroup and sequence variation in mtDNA control region of 61 individuals from the Pas Valley.*

Haplogroup	HVS-I (nps 16024–16383)	Between HVS-I and HVS-II	HVS-II (nps 66–370)	Between HVS-II and HVS-III	HVS-III (nps 438–574)	Sample	Accesion Number
H	–	16519C	263G 309.1C 315.1C	–	–	PAS56	FJ800357
H	–	16519C	146C 263G 315.1C	–	–	PAS35	FJ800336
H	–	16519C	146C 263G 315.1C	–	–	PAS53	FJ800354
H	–	16519C	263G 309.1C 309.2C 315.1C	–	–	PAS03	FJ800304
H	–	16159C	263G 309.1C 309.2C 315.1C	–	–	PAS13	FJ800314
H	16129A	16519C	263G 309.1C 315.1C	–	–	PAS49	FJ800350
H	16233G	–	195C 263G 315.1C	–	d522–523	PAS18	FJ800319
H	16233G	–	195C 263G 315.1C	–	d522–523	PAS19	FJ800320
H	16233G	–	195C 263G 315.1C	–	d522–523	PAS20	FJ800321
H	16233G	–	195C 263G 315.1C	–	d522–523	PAS22	FJ800323
H	16233G	–	195C 263G 315.1C	–	d522–523	PAS44	FJ800345
H	16233G	–	195C 263G 315.1C	–	d522–523	PAS61	FJ800362
H	16148T 16233G	–	195C 263G 315.1C	–	d522–523	PAS26	FJ800327
H	16239T 16242T	16519C	263G 315.1C	–	–	PAS50	FJ800351
V	16298C	–	72C 263G 309.1C 315.1C	–	–	PAS09	FJ800310
V	16298C	–	72C 263G 309.1C 315.1C	–	–	PAS11	FJ800312
V	16298C	–	72C 263G 309.1C 315.1C	–	–	PAS14	FJ800315
V	16298C	–	72C 263G 309.1C 315.1C	–	–	PAS21	FJ800322
V	16298C	–	72C 195C 263G 309.1C 315.1C	–	–	PAS30	FJ800331
V	16298C	–	72C 263G 309.1C 315.1C	–	–	PAS32	FJ800333
V	16298C	–	72C 263G 309.1C 315.1C	–	–	PAS33	FJ800334
V	16298C	–	72C 263G 309.1C 315.1C	–	–	PAS37	FJ800338
V	16298C	–	72C 263G 309.1C 315.1C	–	–	PAS54	FJ800355
V	16298C	16519C	72C 150T 263G 309.1C 309.2C 315.1C	–	–	PAS59	FJ800360
V	16129A 16298C	–	195C 263G 309.1C 309.2C 315.1C	–	523.1CA	PAS55	FJ800356
V	16240G 16298C	–	72C 263G 309.1C 309.2C 315.1C	–	–	PAS28	FJ800329
V	16240G 16298C	–	72C 263G 309.1C 309.2C 315.1C	–	–	PAS46	FJ800347
V	16249C 16298C	–	72C 150T 263G 309.1C 315.1C	–	–	PAS04	FJ800305
V	16249C 16298C	–	72C 150T 263G 309.1C 309.2C 315.1C	–	–	PAS15	FJ800316
U	16288C	16519C	73G 263G 309.1C 315.1C 340T	–	–	PAS02	FJ800303
U	16129A	16519C	73R 263G 309.1C 309.2C 315.1C	–	d522–523	PAS47	FJ800348
U2	16051G 16162G 16264T	16465T 16519C	73G 263G 315.1C	–	d522–523	PAS10	FJ800311
U3	16168T 16343G	16519C	73G 150T 263G 315.1C	–	–	PAS43	FJ800344
U3	16168T 16343G	16519C	73G 150T 263G 315.1C	–	–	PAS45	FJ800346
U4	16356C	16519C	73G 150T 263G 315.1C	–	–	PAS16	FJ800317
U4	16356C	16519C	73G 150T 263G 315.1C	–	–	PAS27	FJ800328
U4	16356C	16519C	73G 150T 263G 315.1C	–	–	PAS38	FJ800339
U4	16356C	16519C	73G 150T 263G 315.1C	–	–	PAS42	FJ800343
U5	16256T 16270T 16320T	16399G	73G 153G 195C 263G 309.1C 315.1C	–	–	PAS60	FJ800361
U5b	16172C 16189C 16270T 16311C	16159C	73G 150T 194T 263G 309.1C 315.1C	–	–	PAS29	FJ800330
K1a	16224C 16311C	16519C	73G 94A 195C 263G 309.1C 315.1C	–	452C 497T	PAS06	FJ800307
K	16224C 16270T 16311C	–	73G 150T 263G 279C 309.1C 315.1C	–	517T	PAS24	FJ800325
K	16224C 16270T 16311C	–	73G 150T 263G 279C 309.1C 315.1C	–	517T	PAS41	FJ800342
T2	16126C 16294T 16296T	16519C	73G 204C 263G 309.1C 315.1C	–	–	PAS17	FJ800318
T2	16126C 16294T 16296T	16519C	73G 204C 263G 309.1C 315.1C	–	–	PAS34	FJ800335
T2	16126C 16294T 16296T	16519C	73G 150T 263G 309.1C 315.1C	–	–	PAS36	FJ800337
T2	16126C 16294T 16296T	16519C	73G 204C 263G 309.1C 315.1C	–	–	PAS39	FJ800340
T2	16126C 16294T 16296T	16519C	73G 204C 263G 309.1C 315.1C	–	–	PAS51	FJ800352
T2	16126C 16294T 16296T	16519C	73G 204C 263G 309.1C 315.1C	–	–	PAS52	FJ800353
T2	16126C 16294T 16296T	16519C	73G 204C 228A 263G 309.1C 315.1C	–	–	PAS40	FJ800341
T2	16126C 16294T 16296T	16519C	73G 204C 253T 263G 309.1C 315.1C	–	–	PAS25	FJ800326

TABLE 1—*(Continued)*

Haplogroup	HVS-I (nps 16024–16383)	Between HVS-I and HVS-II	HVS-II (nps 66–370)	Between HVS-II and HVS-III	HVS-III (nps 438–574)	Sample	Accesion Number
T2b	16126C 16241G 16294T 16296T 16304C	16519C	73G 263G 309.1C 315.1C	–	–	PAS01	FJ800302
T2b	16126C 16241G 16294T 16296T 16304C	16519C	73G 263G 309.1C 315.1C	–	–	PAS05	FJ800306
T2b	16126C 16241G 16294T 16296T 16304C	16519C	73G 263G 309.1C 315.1C	–	–	PAS08	FJ800309
T2b	16126C 16241G 16294T 16296T 16304C	16519C	73G 253T 263G 309.1C 309.2C 315.1C	–	–	PAS58	FJ800359
I1b	16129A 16172C 16223T 16311C	16391A 16519C	73G 199C 203A 204C 250C 263G 315.1C	–	455.1T 573C 573.4C	PAS23	FJ800324
M1	16129A 16183C 16189C 16193.1C 16249C 16311C	16519C	73G 195C 263G 309.1C 309.2C 315.1C	–	489C	PAS07	FJ800308
M1	16129A 16183C 16189C 16193.1C 16249C 16311C	16519C	73G 195C 263G 309.1C 315.1C	–	489C	PAS57	FJ800358
L2c	16223T 16278T 16320C	16390A 16519C	73G 93G 146C 150T 152C 182T 195C 198T 263G 315.1C 325T	–	d522–523	PAS12	FJ800313
L2c	16223T 16278T 16320C	16390A 16519C	73G 93G 146C 150T 152C 182T 195C 198T 263G 315.1C 325T	–	d522–523	PAS31	FJ800332
L2c	16223T 16278T 16320C	16390A 16519C	73G 93G 146C 150T 152C 182T 195C 198T 263G 315.1C 325T	–	d522–523	PAS48	FJ800349

sequence was found to be coincident with the rCRS. Region HVIII proved to be less variable than HVI and HVII: among the 7 haplotypes observed, 3 sequences (4.9%) were unique and 43 sequences (70.5%) were identical to the rCRS.

The HVI-HVII sequences were classified into 30 different haplotypes. The number of different haplotypes did not increase when HVIII was included in the analysis. In the same way, the number of haplotypes remained constant when the entire mtDNA control region (including mtDNA fragments between the hypervariable regions) was considered. The most frequent haplotype was shared by 8 individuals and showed 5 differences with respect to the rCRS (16298C, 72C, 263G, 309.1C, and 315.1C). The haplotype accumulating more nucleotide substitutions relative to the rCRS was characterized by polymorphisms 16223T, 16278T, 16320C, 16390A, 16519C, 73G, 93G, 146C, 150T, 152C, 182T, 195C, 198T, 263G, 315.1C, 325T, and d522–523.

Sequence diversity was first calculated for each subregion separately, and then for the HVI-HVII and HVI-HVII-HVIII complexes and the entire D-loop region (Table 2). The sequence diversity for HVI-HVII haplotypes was estimated to be 0.9596 ± 0.0113. The HVIII region rendered a diversity value of 0.4923 ± 0.0603, which is below the diversity estimates for other European and Asian populations, namely, Bologna (0.590) (4) and Japan (0.726) (5), and similar to Finnish (0.5044 ± 0.0439; calculated from data by Finnila et al.) (20). Overall, the sequence diversity in the Pas Valley was among the lowest in Europe (see for instance Alfonso-Sánchez et al., 2008 for comparative diversity data).

In explaining the low diversity found for the mitochondrial genome control region of the Pas Valley population, the following points must be considered: (i) the existence of high endogamy and/or inbreeding levels in the studied area (8,21), intimately related to the geography of the territory (with predominance of a very complex orography), and thereby to the geography of peopling, characterized mainly by dispersed population and moderate geographic distances between relatively isolated population nuclei. In addition, the Pas Valley region is still predominantly rural, and the local population is characterized by a strong adherence to traditional sociocultural mores, where consanguineous marriages have been an important component of the marital structure of the population, as demonstrated in other northern Spanish regions (22,23), and (ii) the potential impact of severe genetic drift episodes resulting mainly from population bottlenecks, bearing in mind the small demographic size of the population nuclei in the area, which would result in the predominance of a few female genetic lineages, thereby reducing the overall mtDNA diversity.

The random match probability for the complex HVI-HVII, the most commonly analyzed estimate in human mtDNA variation studies, indicated that the probability of two randomly selected individuals from the Pas Valley population having identical mtDNA types is 5.62%. This probability remained constant for both the HVI-HVII-HVIII complex and the entire mtDNA control region. Logically, the genetic isolation of the Pas Valley was clearly reflected in the comparatively high value of random match probability. Thus, we confirmed that the analysis of the entire mitochondrial DNA control region in highly isolated populations shows only a limited usefulness for identification purposes.

Haplogroup Classification

All the main European-specific haplogroups were represented in our sample: H (23.0%), V (24.6%), U (18.0%), K (4.9%), T (19.7%), and I (1.6%). In all, these haplogroups accounted for 91.8% of mtDNA lineage diversity in the Pas Valley collection.

TABLE 2—*Diversity and forensic parameters computed from entire mtDNA control region analysis in the population from the Pas Valley (Spain).*

	HVS-I	HVS-II	HVS-III	HVS-I + HVS-II	HVS-I + HVS-II + HVS-III	Control Region
Sequence diversity	0.9350 ± 0.0145	0.9295 ± 0.0145	0.4923 ± 0.0673	0.9596 ± 0.0113	0.9596 ± 0.0113	0.9596 ± 0.0113
Mean number of pairwise difference	3.6765 ± 1.8856	2.8891 ± 1.5391	1.1126 ± 0.7368	6.5654 ± 3.1455	7.5339 ± 3.5660	7.7601 ± 3.6642
Random match probability	8.04%	8.58%	51.65%	5.62%	5.62%	5.62%
Nucleotide diversity	0.0102 ± 0.0058	0.0094 ± 0.0056	0.0079 ± 0.0058	0.0099 ± 0.0058	0.0093 ± 0.0049	0.0069 ± 0.0036
Substitutions	30	21	5	51	56	61
Indels	0	0	5	0	5	5
Polymorphic sites	30	21	10	51	61	66
Haplotypes	22	21	7	30	30	30
Insertion in the poly-C stretches	1	3	4	4	8	8

On the other hand, we found several non-European sequences (M1: 3.3% and L2c: 4.9%). These findings coincide with the haplogroup distribution observed in a previous study focused on mitochondrial genome diversity in different Cantabrian populations (9). The two mtDNA sequences classified into haplogroup M1 showed a back mutation at position 16223. This polymorphism was previously described by Maca-Meyer et al. (9) in an individual from the Pas Valley, assigned to haplogroup M1 by means of restriction fragment length polymorphism (RFLP) analysis. In our case, the study of the entire mtDNA control region including nps 195 and 489, which are diagnostic positions for M1 (24), permitted reliable classification of these individuals into clade M1. Concerning haplogroups K and I, analysis of segment HVIII made it possible to refine the classification into subhaplogroups K1a and I1b, respectively (25).

Some studies focused on the characterization of the polymorphism of segment HVIII have suggested that this approach could be particularly useful for forensic purposes. Thus, for instance, Lutz et al. (3) found that "about one-fifth of all sequences that were identical in the HVI and HVII regions could be distinguished by additional analysis of the HVIII region" in a sample comprising Germans, Austrians, and Swiss. The usefulness of HVIII in forensic applications was further confirmed by Bini et al. (4) and Mabuchi et al. (5). In contrast, the results of this study indicate that even the analysis of the entire mtDNA control region may have important limitations for use in forensic casework when dealing with human isolates: none of the 44 individuals who exhibited identical HVI-HVII haplotypes could be further differentiated by the analysis of segment HVIII. Nevertheless, our analysis of the entire mtDNA control region proved to be useful to determine the ancestry of the samples examined, by contributing to the confirmation, and on occasion, even to the refinement of the haplogroup assignment. In fact, a precise haplogroup classification is an issue of increasing importance when mtDNA analysis is applied to forensic casework (26–28). Obviously, knowing the ancestry (geographic origin and/or ethnicity) of a sample may be the important first step in forensic cases for which only a very limited amount of DNA is available (DNA vestiges). In these latter cases, implementing the analysis of the entire mtDNA control region as a routine procedure in forensic investigation would be advisable.

In all, we can conclude that in isolated human groups, where mitochondrial genome diversity may be substantially low, although the improvement of the power of discrimination between sequences by analysis of the entire mtDNA control region seems to be somewhat limited, this procedure might constitute a very useful tool to carry out a reliable, unambiguous haplogroup identification.

Acknowledgments

The research was supported by the University of the Basque Country (GIU 05/51) and the Basque Government (IT-424). We also thank the Servicios Generales de Investigación (SGIker) of the University of the Basque Country for their support. In addition, the authors gratefully acknowledge all voluntary donors from the Pas Valley who made this study possible.

References

1. Budowle B, Allard MW, Wilson MR, Chakraborty R. Forensics and mitochondrial DNA: applications, debates, and foundations. Annu Rev Genomics Hum Genet 2003;4:119–41.
2. Butler JM. Forensic DNA typing: biology, technology, and tenetics of STR markers, 2nd edn. New York, NY: Elsevier Academic Press, 2005.
3. Lutz S, Wittig H, Weisser HJ, Heizmann J, Junge A, Dimo-Simonin N, et al. Is it possible to differentiate mtDNA by means of HVIII in samples that cannot be distinguished by sequencing the HVI and HVII regions? Forensic Sci Int 2000;113:97–101.
4. Bini C, Ceccardi S, Luiselli D, Ferri G, Pelotti S, Colalongo C, et al. Different informativeness of the three hypervariable mitochondrial DNA regions in the population of Bologna (Italy). Forensic Sci Int 2003;135:48–52.
5. Mabuchi T, Susukida R, Kido A, Oya M. Typing the 1.1 kb control region of human mitochondrial DNA in Japanese individuals. J Forensic Sci 2007;52:355–63.
6. Alfonso-Sánchez MA, Cardoso S, Martinez-Bouzas C, Pena JA, Herrera RJ, Castro A, et al. Mitochondrial DNA haplogroup diversity in Basques: a reassessment based on HVI and HVII polymorphisms. Am J Hum Biol 2008;20(2):154–64.
7. Cardoso S. Diversidad del Genoma Mitocondrial en Poblaciones Autóctonas de la Cornisa Cantábrica: el Repoblamiento Postglacial de Europa [PhD Thesis]. Spain: University of the Basque Country, 2008.
8. Gómez P, Arminio ML. Matrimonio entre pasiegos. I. Endogamia/exogamia y consanguinidad en la población de los Montes de Pas, 1880–1979. Publ Inst Etnol Folk "Hoyos Sainz" 2001;XV:79–105.
9. Maca-Meyer N, Sanchez-Velasco P, Flores C, Larruga JM, Gonzalez AM, Oterino A, et al. Y chromosome and mitochondrial DNA characterization of Pasiegos, a human isolate from Cantabria (Spain). Ann Hum Genet 2003;4:329–39.
10. Brion M, Quintans B, Zarrabeitia M, Gonzalez-Neira A, Salas A, Lareu V, et al. Micro-geographical differentiation in Northern Iberia revealed by Y-chromosomal DNA analysis. Gene 2004;329:17–25.
11. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet 1999;23(2):147.
12. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 1997;25(24):4876–82.
13. Excoffier L, Laval G, Schneider S. Arlequin ver. 3.0: an integrated software package for population genetics data analysis. Evol Bioinform Online 2005;1:47–50.

14. Nei M. Molecular evolutionary genetics. New York, NY, USA: Columbia University Press, 1987.
15. Tajima F. Evolutionary relationship of DNA sequences in finite populations. Genetics 1983;105(2):437–60.
16. Stoneking M, Hedgecock D, Higuchi RG, Vigilant L, Erlich HA. Populations variation of human mtDNA control region sequences detected by enzymatic amplification and sequence-specific oligonucleotide probes. Am J Hum Genet 1981;48:370–82.
17. Bär W, Brinkmann B, Budowle B, Carracedo A, Gill P, Holland M, et al. DNA Commission of the International Society for Forensic Genetics: guidelines for mitochondrial DNA typing. Int J Legal Med 2000;113(4):193–6.
18. Lutz S, Weisser HJ, Heizmann J, Pollak S. Location and frequency of polymorphic positions in the mtDNA control region of individuals from Germany. Int J Legal Med 1998;111(2):67–77.
19. Palanichamy MG, Sun C, Agrawal S, Bandelt HJ, Kong QP, Khan F, et al. Phylogeny of mitochondrial DNA macrohaplogroup N in India, based on complete sequencing: implications for the peopling of South Asia. Am J Hum Genet 2004;75(6):966–78.
20. Finnila S, Lehtonen MS, Majamaa K. Phylogenetic network for European mtDNA. Am J Hum Genet 2001;68(6):1475–84.
21. Fuster V, Colantonio SE. Consanguinity in Spain: socioeconomic, demographic, and geographic influences. Hum Biol 2002;74:301–15.
22. Varela TA, Lodeiro R, Fariña J. Evolution of consanguinity in the Archbishopric of Santiago de Compostela (Spain) during 1900–1979. Hum Biol 1997;69:517–31.
23. Alfonso-Sánchez MA, Aresti U, Peña JA, Calderón R. Inbreeding levels and consanguinity structure in the Basque province of Guipúzcoa (1862–1980). Am J Phys Anthropol 2005;127:240–52.
24. Olivieri A, Achilli A, Pala M, Battaglia V, Fornarino S, Al-Zahery N, et al. The mtDNA legacy of the Levantine early Upper Palaeolithic in Africa. Science 2006;314(5806):1767–70.
25. van Oven M, Kayser M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum Mutat 2009;30(2): E386–94.
26. Grignani P, Peloso G, Achilli A, Turchi C, Tagliabracci A, Alù M, et al. Subtyping mtDNA haplogroup H by SNaPshot minisequencing and its application in forensic individual identification. Int J Legal Med 2006;120(3):151–6.
27. Asari M, Umetsu K, Adachi N, Azumi J, Shimizu K, Shiono H. Utility of haplogroup determination for forensic mtDNA analysis in the Japanese population. Leg Med 2007;9(5):237–40.
28. Nelson TM, Just RS, Loreille O, Schanfield MS, Podini D. Development of a multiplex single base extension assay for mitochondrial DNA haplogroup typing. Croat Med J 2007;48(4):460–72.

Additional information and reprint requests:
Marian M. de Pancorbo, Ph.D.
Professor
BIOMICS Research Group
Centro de Investigación y Estudios Avanzados ''Lucio Lascaray''
Universidad del País Vaso
Miguel de Unamuno 3, 01006
Spain
E-mail: marianpancorbo@gmail.com

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01423.x
Available online at: interscience.wiley.com

PAPER

CRIMINALISTICS

Eun J. Lee,[1] Ph.D.; In K. Hwang,[1] M.S.; Nam Y. Kim,[2] M.D.; Kyung L. Lee,[1] Ph.D.; Myun S. Han,[2] Ph.D.; Yang H. Lee,[2] Ph.D.; Mu Y. Kim,[3] Ph.D.; and Moon S. Yang,[3] Ph.D.

An Assessment of the Utility of Universal and Specific Genetic Markers for Opium Poppy Identification

ABSTRACT: The proper identification of illicit plants such as *Papaver somniferum* L (opium poppy) is important for law enforcement agencies. The identification of opium poppy was presently tested using 10 genetic markers that are universal for all plants or specific to a few poppy plants. The genetic distances of universal markers such as nuclear internal transcribed spacer (ITS), 18S rRNA, plastid *rbc*L, and *trn*L-*trn*F intergenic spacer (IGS) of 14 species included in the Papaveraceae and Fumariaceae family were acquired by sequence comparisons. Both the ITS region and *trn*L-*trn*F IGS showed high levels of interspecific divergence. Six *Papaver* genera-specific markers were developed from coding regions involved in morphine biosynthesis. Three markers (TYDC, NCS, and BBE) produced amplicons only in opium poppy, providing a presence/absence test for opium poppy, while three additional markers (CYP80B1, SAT, and COR) were genus specific. These 10 markers might be useful for the forensic DNA analysis of opium poppy.

KEYWORDS: forensic science, *Papaver somniferum* L, genetic markers, morphine biosynthesis, profiling, narcotic control

Papaver somniferum L is one of the most important medicinal plants. The possession of any part of the opium poppy other than the seeds and cultivation of the plant is outlawed in South Korea and in other countries. The seeds of opium poppy are hard to distinguish visually from other poppy seeds. Currently, the identification of opium poppy depends mostly on the chemical analysis (1). Extracting opiate from poppy seed is particularly troublesome because poppy seed is an excellent emulsifier. Marijuana (*Cannabis sativa* L), another important medicinal plant, can be identified using universal (2) or specific (3) polymerase chain reaction (PCR) primer sets. Knowledge of the genetic structure of the opium poppy would facilitate the control of drug trafficking by enabling detection from tiny fragments (or seeds) of a plant.

In seeking genetic means of poppy identification, the sequence characteristics of genetic loci were used. Universal markers of these genetic loci included the internal transcribed spacer (ITS) (4) and 18S ribosomal RNA (rRNA) (5) located in the genomic nucleolar organizing region; the gene encoding the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (*rbc*L), which catalyzes the carbon fixation for energy metabolism (6,7); and the *trn*L-*trn*F intergenic spacer (IGS) (8,9) located in the chloroplast DNA. These are currently used for the phylogenetic studies of various plants. Tyrosine/dopa decarboxylase (TYDC) (10), S-norcoclaurine synthase 1 (NCS) (11), (S)-N-methylcoclaurine 3′-hydroxylase (CYP80B1) (12), berberine bridge enzyme (BBE) (13), salutaridinol 7-O-acetyltransferase (SAT) (14), and codeinone reductase (COR) (15), which participate in the benzylisoquinoline alkaloid biosynthetic pathway, were also utilized as *Papaver* genera-specific markers accessed by presence or absence of each amplicon.

Using universal markers, we aimed to evaluate the phylogenetic relationship of opium poppy with similar poppy plants and generate sequence information of poppy plants for the application to the future forensic cases. Also, we tried to discriminate opium poppy from other poppy plants using a presence/absence test of *Papaver* genera-specific markers such as TYDC, NCS, CYP80B1, BBE, SAT, and COR.

Materials and Methods

Plant Material and Isolation of Total DNA

The plant specimens of 14 species used in this study are summarized in Table 1. DNA from fresh leaves, fresh seedlings of purchased seeds, and dried leaves were utilized. Plant DNA was extracted from a selection of tissues using the DNeasy Plant Mini Kit (QIAGEN, Valencia, CA). The isolated DNA was quantified using agarose gel electrophoresis and stored at −20°C until use. To assess species specificity of *Papaver* genera markers, DNA of rice (*Oryza sativa*), Japanese hop (*Humulus japonicus*), honeysuckle (*Lonicera japonica*), bamboo (*Phyllostachys* species), cypress (*Chamaecyparis obtusa* Endl.), clover (*Trifolium repens*), and 9947A human female DNA were tested. To examine the appropriate tissue samples for DNA source aiming at the future drug seizures, DNA from fresh and air-dried leaf, fresh stem and root, and seeds of opium poppies were assessed.

[1]DNA Analysis Sector, Western District Office of National Institute of Scientific Investigation, 111 Daeduk-Ri, Seosam-Myun, Jangsung-Gun, Chonnam, 515-822, South Korea.
[2]DNA Analysis Sector, National Institute of Scientific Investigation, 331-1 Shinwol-7Dong, Yangchun-Ku, Seoul, 158-707, South Korea.
[3]Department of Biological Sciences, Chonbuk National University, 664-14 Dukjin-Dong, Dukjin-Gu, Chonju City, Chonbuk, 561-756, South Korea.

Received 19 Jan. 2009; and in revised form 23 June 2009; accepted 27 June 2009.

TABLE 1—*Taxa used in this study, their collected source, and their NCBI accession numbers.*

							GenBank Accession No.							
Family	Genus	Species	Common Name	N*	Abb†	Source‡	ITS	rRNA	rbcL	trnLF	NCS	CYP80B1	SAT	COR
Papaveraceae	*Chelidonium*	*Chelidonium majus*	Greater celandine	1	C.MAJU, CM	Cbnu	DQ912878	DQ912864	DQ912892	DQ912906	–	–	–	–
	Hylomecon	*Hylomecon vernalis*	Forest poppy	1	H.VERN, HV	Cbnu	DQ912877	DQ912863	DQ912891	DQ912907	–	–	–	–
		Hylomecon hylomecoides	Mae-mi-kkot§	1	H.HYLO, HH	Cbnu	DQ912876	DQ912862	DQ912890	DQ912908	–	–	–	–
	Papaver	*Papaver radicatum*	Arctic poppy§	1	P.RADI, PRA	Cbnu	DQ912879	DQ912865	DQ912893	DQ912909	–	–	–	–
		Papaver bracteatum	Oriental poppy	1	P.BRAC, PB	TM	DQ912881	DQ912870	DQ912897	DQ912911	–	FJ596171	FJ200355 FJ200356	FJ596162 FJ596163
		Papaver nudicaule	Iceland poppy	1	P.NUDI, PN	TM	DQ912885	DQ912871	DQ912898	DQ912912	–	FJ596172	-	FJ596164 FJ596165
		Papaver orientale	Oriental poppy	6	P.ORIE, PO	TM	DQ912882	DQ912872	DQ912899	DQ912913	–	FJ596173	FJ200357 FJ200358	FJ596166 FJ596167
		Papaver somniferum L	Opium poppy	3	P.SOMN, PS	Chonnam, TM	DQ912880	DQ912867	DQ912894	DQ912910	FJ200359	FJ596170	FJ200353 FJ200354	FJ596158 FJ596159 FJ596160 FJ596161
		Papaver rhoeas	Corn poppy	2	P.RHOE, PRH	TM	DQ912886 FJ469600	DQ912866	DQ912900	DQ912914	–	FJ596174	–	FJ596168 FJ596169
	Eschscholzia	*Eschscholzia californica*	California poppy	1	E.CALI, EC	TM	DQ912883	DQ912868	DQ912895	DQ912904	–	–	–	–
		Eschscholzia mexicana	Mexican gold poppy	1	E.MEXI, EM	TM	DQ912884	DQ912869	DQ912996	DQ912905	–	–	–	–
Fumariaceae	*Corydalis*	*Corydalis ambigua*	Yan Hu Suo	2	C.AMBI, CA	Chonnam, eshop	DQ912888 FJ469597 FJ469598	DQ912873	DQ912902	DQ912916	–	–	–	–
		Corydalis incisa Pers	–	1	C.INCI, CI	Chonnam	DQ912889 FJ469599	DQ912874	DQ912903	DQ912917	–	–	–	–
	Dicentra	*Dicentra spectabilis*	Bleeding heart	1	D.SPEC, DS	Chonnam	DQ912887	DQ912875	DQ912901	DQ912915	–	–	–	–

*N represents individual number of sample.

†Abb.: abbreviation for each taxa used in this study.

‡Source means collected or purchased sites of plant sample (Cbnu, Chonbuk National University; TM, Thompson and Morgan (UK) Limited; Chonnam, Chonnam province in South Korea; eshop, purchased from Internet shopping mall in South Korea).

§Species indigenous to Korea.

ITS, internal transcribed spacer.

TABLE 2—*Primer sequences used in this study.*

Markers	Primer	Oligo Sequences	Gene Accession Number	Reference
ITS	ITS 2	5′-GCTGCGTTCTTCATCGATGC-3′	Intergenic transcribed spacer	4
	ITS 3	5′-GCATCGATGAAGAACGCAGC-3′		
	ITS 4	5′-TCCTCCGCTTATTGATATGC-3′		
	ITS 5	5′-GGAAGTAAAAGTCGTAACAAGG-3′		
18S rRNA	F	5′-TGCAGTTAAAAAGCTCGTAG-3′	rRNA of large subunit of ribosome	5,16
	R	5′-GGTTGAGACTAGGACGGTATCTG-3′		LEJ
*rbc*L	F	5′-AATTCATGAGTTGTAGGGAGGGA-3′	RuBisCo	LEJ
	R	5′-AATTCGCAGATCCTCCAGACGT-3′		
*trn*L-*trn*F	F	5′-AAAATCGTGAAGGTTCAAGTC-3′	Chloroplast intergenic spacer	8,9
	R	5′-GATTTGAACTGGTGACACGAG-3′		
TYDC	F	5′-CCGTCTCAACCAGATCAA-3′	*tydc*3	LEJ
	R	5′-TTCACCGCATACCTTTACG-3′	AF025431	
NCS	F	5′-ATGTCTAAGTTGATCACGAC-3′	*ncs*1	LEJ
	R	5′-TTAGAGTGGAACACCAGCAA-3′	AY860500	
CYP80B1	F	5′-TGTGATCACTACTTTCTTA-3′	*cyp80b*1	LEJ
	R	5′-CATAGCCTTACTTTCTCTAAC-3′	AF191772	
BBE	F	5′-CCTGCGGGTTATTAAAACC-3′	*bbe*	LEJ
	R	5′-TGTTGCCCCAGATTCAAC-3′	AF025430	
SAT	F	5′-ATGGCAACAATGTATAGTGC-3′	*sal*AT	LEJ
	R	5′-TCAAATCAATTCAAGGATTT-3′	AF339913	
	2R	5′-ATCTGTAACGGCAGCAGTTGAAC-3′		
COR	F	5′-ATGGAGAGTAATGGTGTACC-3′	*cor*1-1	LEJ
	R	5′-TCAATCCTTCTCATCCCAGA-3′	AF108432	

"F," forward primer; "R," reverse primer; ITS, internal transcribed spacer; LEJ, primer is designed by author, Eun-jung Lee.

Molecular Markers

The primer sequences used in this study are shown in Table 2. The universal primers for the botanical species identification were used to amplify the ITS (4) and the *trn*L-*trn*F IGS (9). The primer set for 5′ partial *rbc*L (fragment length of 446 bp, −26 to 420) and the reverse primer for 5′ partial 18S rRNA (fragment length of approximately 412 bp) were presently designed to maximize the PCR outcome from very small sample quantities. The Angio 1F forward primer for 18S rRNA has been previously described (16). *Papaver* genera-specific primer pairs were designed on the basis of previously National Center for Biotechnology Information (NCBI)-registered TYDC, NCS, CYP80B1, BBE, SAT, and COR (Table 2). PCR primer sequences of TYDC and BBE were based on genomic DNA and those of NCS, CYP80B1, SAT, and COR on complementary DNA (cDNA). In the case of cDNA, forward primers were designed to include the adjacent start codon sequence, while the reverse primers included the stop codon (except CYP80B1). TYDC primer set was designed on the basis of the alignment of the tyrosine/dopa decarboxylase (*tydc*) alleles (1, 2, 3, 4, 6, 7, 8, and 9) of *P. somniferum*, of *Arabidopsis thaliana* (AT2G20340 mRNA), and of *Oryza sativa* (japonica cultivar-group, OSJNBa0050N08.21). Among the several COR alleles, we chose *cor1-1*. GeneFisher (Bielefeld University bioinformatics server, http://bibiserv.techfak.uni-bielefeld.de/genefisher2/submission.html) was used to design BBE primers for PCR. To increase the opium-specificity of SAT, a second reverse primer (SAT-2R) bound to the 5′ upstream region of the first reverse primers was designed.

PCR Amplification and DNA Sequencing

All amplification were performed in 50 μL of reaction medium containing 20 pmol of forward and reverse primers synthesized by BIONEER Corporation (Daejon, South Korea), and about 10–100 ng of genomic DNA. The PCR mixture was incubated for 11 min at 95°C, 28 cycles of 1 min at 95°C, 1 min at 50°C, 1 min at 72°C, and 7 min at 72°C, using an ABI 9700 apparatus (Applied Biosystems, Foster City, CA). A QIAquick PCR purification kit (QIAGEN) was used for the primer removal. An ABI PRISM Big-Dye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) and the DyeEX™ 2.0 spin kit (QIAGEN) were used for DNA sequencing. Automated DNA sequencing was conducted using an ABI 310 sequencer (Applied Biosystems).

Vector Cloning

When direct sequencing was not possible because of amplicon length polymorphism in three species (ITSs of *Papaver rhoeas*, *Corydalis ambigua*, and *C. incisa*) and large amplicon sizes (TYDC, NCS, SAT, and COR), the PCR products were cloned into a pGEM-T vector (Promega, Madison, WI). The plasmid DNA harboring insert was isolated using a Wizard Miniprep kit (Promega). The isolated plasmids were sequenced using M13 universal primers or by marker primer itself.

Alignments and Tree Construction

The sequences of the molecular markers determined from the 14 species were arranged using Molecular Evolution Genetics Analysis (MEGA) software (17). The sequences (Table 1) were registered in NCBI (18). If putative heterozygosity appeared at a specific nucleotide position, a higher peak was selected for the decision of sequence information. Nucleotides with the same peak heights were observed at one site, then they were designated as "N." In species having multiple ITS sequences such as *P. rhoeas*, *C. ambigua*, and *C. incisa*, DQ912886, DQ912888, and DQ912889, respectively, were selected for tree construction (Fig. 1, Table 1). Neighbor-joining trees with bootstrapped value were constructed using MEGA software.

Results

Sequence Comparisons Using Universal Genetic Markers

Samples of the 14 species were amplified and sequenced using the primer sets of ITS, 18S rRNA, *rbc*L, and *trn*L-*trn*F IGS.

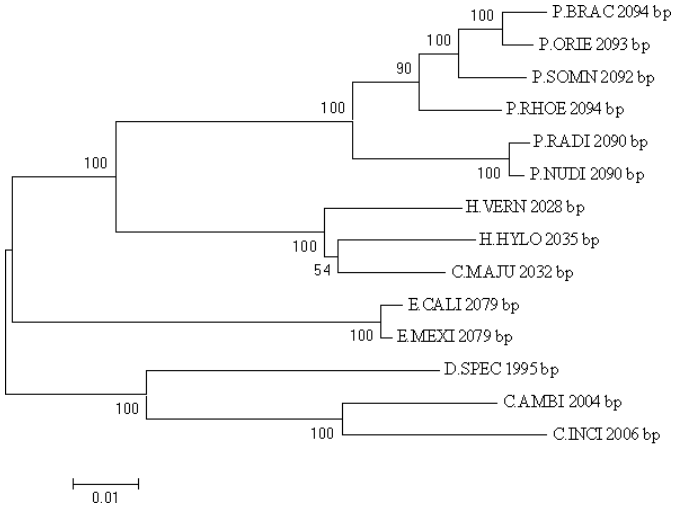


FIG. 1—*Bootstrapped neighbor-joining tree of the ITS + 18S rRNA* + rbc*L* + trn*L*-trn*F IGS sequences from plant species used in this study. Total nucleotide lengths of each species are shown in the vicinity of taxa names.*

Table 1 shows the GenBank accession numbers registered in NCBI.

Identification Using Universal ITS and 18S rRNA in nrDNA

The entire ITS region ranged from 673 bp in *C. ambigua* to 779 bp in *P. bracteatum*. The length ranges of ITS1 and ITS2 spacers were 176–251 bp and 205–254 bp, respectively. Based on the entire ITS sequence, the most evolutionarily conserved species of opium poppy was *P. rhoeas*, and the least conserved was *C. incisa* (Table 3). Each genus had regular size patterns of ITS amplicons. The mean length of ITS region from the *Papaver*, *Hylomecon*, and *Eschscholzia* genera were 776 bp, 710 bp, and 731 bp, respectively.

Intraspecies length variations and sequence diversities were detected in the ITS regions of three species (*P. rhoeas*, *C. ambigua*, and *C. incisa*). ITS clones of these species were acquired and revealed 1 bp length differences, and several point mutations were detected in one individual. In the ITS region of *P. rhoeas*, sequences of 778 bp and 779 bp were cloned; they had 22 transitions/transversions and three indels between each other. In the ITS region of *C. ambigua* (leaf part), one clone of 673 bp and two clones of 674 bp were secured. The two 674 bp clones had nine transitions/transversions between each other. Some ITS clones of *C. ambigua* (a bulbous part) purchased from an Internet shopping mall showed high homologies with *Corydalis flavula* and several fungi. But *C. ambigua* (leaf part) gathering from Chonnam province showed no fungal sequences. The degree of fungal contamination might depend on the organ from which genomic DNA were isolated. In the ITS region of *C. incisa*, two clones of 678 bp were acquired. They had one transversion and three transitions between each other. In cases of ITS of *P. rhoeas*, *C. ambigua*, and *C. incisa*, it is worth observing more clones to deduce their evolutionary histories.

The 18S rRNA region was 412–415 bp in length. Even though there were intra-specific variants or fungal contamination in the ITS marker, which was located close to the 18S rRNA, such length variations or fungal sequences were not detected in the 18S rRNA using the direct sequencing method. It was not necessary to clone the 18S rRNA region. Sequence variations (nucleotide substitutions and indels) between taxa were frequently found toward the 5′ end of the DNA sequence. One G insertion was detected at position 56 bp in Mae-mi-kkot, and one T insertion was detected at position 57 bp in Mae-mi-kkot and in greater celandine. One C deletion was detected at position 52 bp in California poppy, and one G deletion was detected at position 56 bp in opium poppy.

Identification Using Universal rbc*L and* trn*L*-trn*F IGS in Chloroplast Genome*

The 5′ partial *rbc*L regions (446 bp) in cpDNA from 14 species were successfully amplified and sequenced using the direct sequencing method. The *rbc*L nucleotide and amino acid sequences of Mae-mi-kkot and greater celandine were the same and differed from that of *H. vernalis* at one site (ACT → AAT at 94 bp: Thr → Asn). The same nucleotide and amino acid sequences were observed in *P. bracteatum* versus *P. orientale* and in *P. nudicaule* versus *P. radicatum*. There were three nucleotide alterations at position 153 bp, 378 bp, and 393 bp between *P. somniferum* and *P. rhoeas*, but no change was observed in their translated amino

TABLE 3—*Evolutionary distances calculated from the entire ITS sequences of 14 species.*

	HH	HV	CM	PRA	PS	PB	PO	PN	PRH	EC	EM	CA	CI	DS
HH														
HV	0.079													
CM	0.058	0.058												
PRA	0.158	0.161	0.154											
PS	0.152	0.154	0.139	0.064										
PB	0.167	0.162	0.150	0.071	0.026									
PO	0.158	0.160	0.141	0.065	0.026	0.024								
PN	0.154	0.163	0.150	0.010	0.062	0.069	0.064							
PRH	0.156	0.148	0.145	0.058	0.021	0.028	0.024	0.060						
EC	0.181	0.168	0.168	0.167	0.169	0.166	0.162	0.169	0.162					
EM	0.184	0.171	0.170	0.169	0.171	0.171	0.164	0.171	0.164	0.003				
CA	0.215	0.195	0.203	0.200	0.204	0.195	0.195	0.198	0.200	0.154	0.154			
CI	0.225	0.188	0.204	0.217	0.207	0.202	0.199	0.219	0.200	0.169	0.171	0.078		
DS	0.212	0.195	0.191	0.191	0.171	0.166	0.168	0.182	0.173	0.190	0.190	0.164	0.177	

The overall mean distance is 0.148.

ITS, internal transcribed spacer; HH, *Hylomecon hylomecoides*; HV, *H. vernalis*; CM, *Chelidonium maju*; PRA, *Papaver radicatum*; PS, *P. somniferum*; PB, *P. bracteatum*; PO, *P. orientale*; PN, *P. nudicaule*; PRH, *P. rhoeas*; EC, *Eschscholzia californica*; EM, *E. mexicana*; CA, *Corydalis ambigua*; CI, *C. incisa*; DS, *Dicentra spectabilis*.

TABLE 4—*Average values analyzed by MEGA software from the universal marker sequences of 14 species.*

	ITS	18S rRNA	*rbc*L	*trn*L-F IGS	Combined
Mean length (bp)	734.4	412.8	446	464.6	2057.8
Mean G+C content (%)	58.3	49.5	43.6	39.3	47.7
Composition distance	0.540	0.073	0.040	0.162	0.540
Disparity index	0.428	0.052	0.014	0.068	0.464
Overall mean distance	0.148	0.022	0.032	0.139	0.094

"Combined" means that internal transcribed spacer (ITS), 18S rRNA, *rbc*L, and *trn*L-*trn*F IGS were joined together to form a single characteristic of each taxa.

acids. The amino acid sequences of *rbc*L from *C. ambigua* and *C. incisa* versus *D. spectabilis* revealed only one alteration among them, but their nucleotide sequences had seven or 15 alterations, respectively.

The *trn*L-*trn*F IGS region was 456–490 bp in length. Each genus had a regular size pattern like the ITS region. The mean length of the *trn*L-*trn*F IGS from *Papaver*, *Hylomecon*, and *Eschscholzia* genera were 457 bp, 462 bp, and 490 bp, respectively. In a comparison of six randomly chosen individuals of *P. orientale*, intraspecies length variations and sequence diversities were not detected in the *trn*L-*trn*F IGS region.

The phylogenetic relationships derived from combined sequences of ITS, 18S rRNA, *rbc*L, and *trn*L-*trn*F IGS of each of the 14 species were assignable to one tree (Fig. 1). The bootstrap values for individual clades based on 1000 replicates ranged from 90% to 100%, with the exception of the clade of the common ancestor of Mae-mi-kkot and greater celandine, which was 54%.

Table 4 shows the composition distance, disparity index, and overall mean distance that demonstrates the phylogenetic distances of 14 species calculated from nucleotide characteristics of each universal marker and their combined characteristics obtained by adding up ITS, 18S rRNA, *rbc*L, and *trn*L-*trn*F IGS sequences of each species (19). Composition distance, which is a measure of the difference in nucleotide composition for a given pair of sequences, was highest in comparisons of ITS sequences and of combined ones and was lowest in comparisons of *rbc*L sequences.

Presence/Absence Test Using Papaver Genera-Specific Markers

The NCBI-registered sequences of TYDC, NCS, CYP80B1, BBE, SAT, and COR from *P. somniferum* were used as the template for designing of each primer pair (10–15, Table 2). These six markers were tested for the profiling of six *Papaver* genera species and two *Eschscholzia* genera (Table 5).

Discrimination of Poppy Plants Using TYDC, NCS, and BBE Regions

The TYDC primers produced a 1371-bp amplicon from only opium poppy of the eight poppy plants tested. The 560-bp-sized 5′ partial region of the TYDC bacterial clone showed 99.8% homology with AF025431 (G641A).

NCS primer pairs made an 856-bp-sized amplicon from opium poppy. Some species like *Eschscholzia californica* (EC), *Eschscholzia mexicana* (EM), *Papaver bracteatum* (PB), and *Papaver orientale* (PO) showed nonspecific faint bands. Our NCS had two point mutations in comparison with AY860500 (A278T, A456G in exon region). Based on comparison with *ncs*1 (AY860500) in NCBI database, one 160 bp intron was detected between 325 bp and 326 bp.

Primer pairs of BBE marker made 516-bp-sized amplicon that corresponded to 2579–3074 bp of AF025430. Our BBE sequence was identical to AF025430. After amplification reaction with BBE primer pairs, a PCR band was detected only in the opium poppy, similar to that found using the TYDC and NCS markers. But some species such as *P. bracteatum* (PB), *P. nudicaule* (PN), *P. radicatum* (PRA), and *P. rhoeas* (PRH) showed nonspecific faint bands that disappeared when annealing temperature was raised to 55°C.

Discrimination of Poppy Plants Using CYP80B1, SAT, and COR Regions

An amplified product for CYP80B1 was detected from *P. bracteatum*, *P. nudicaule*, *P. orientale*, *P. rhoeas*, and *P. somniferum* (Table 5). The length of the amplified product from all five species was 418 bp at the 5′ region. There was one alteration at 297 bp between *P. bracteatum* and *P. orientale*, but no change was observed in their translated amino acids. Based on the nucleotide sequences of CYP80B1, the opium poppy was most closely related to *P. orientale* and least related to *P. nudicaule*.

The forward and reverse primers for SAT bound near the ATG start codon and stop codon, respectively, and produced amplified fragments from oriental poppies and opium poppy (Table 5). The SAT nucleotide sequences of two clones from each species were determined (Table 1). No introns were found in the SAT region of opium poppy; they were all of exons. One nucleotide substitution (A974C) was detected in two opium SAT clones (accession numbers FJ200353 and FJ200354) compared to AF339913 in the NCBI database. *P. bracteatum* produced two amplified fragments of the SAT marker; one (accession number FJ200355) similar to the fragment from *P. somniferum* and the other (accession number FJ200356) similar to that of *P. orientale*. SAT sequences of oriental poppies had seven indels that consist of three or six nucleotides

TABLE 5—*Comparisons of the presence or absence of PCR products in Papaver genera-specific markers in poppy plants.*

Markers	Amplicon Size in PS (bp)	EC	EM	PB	PN	PO	PRA	PRH	PS
TYDC	1371	−	−	−	−	−	−	−	+
CYP80B1	418	−	−	+	+	+	−	+	+
NCS	856	−	−	−	−	−	−	−	+
BBE	516	−	−	−	−	−	−	−	+
SAT	1425	−	−	+	−	+	−	−	+
SAT2	588	−	−	−	−	−	−	−	+
COR	1324	−	−	−	+	−	−	+	+
	1643	−	−	+	−	+	−	−	+

Gene size is based on the sequences of opium poppy.

"+" or "−" means the presence or absence of amplicons.

EC, *Eschscholzia californica*; EM, *E. mexicana*; PB, *P. bracteatum*; PN, *P. nudicaule*; PO, *P. orientale*; PRA, *P. radicatum*; PRH, *P. rhoeas*; PS, *P. somniferum*.

per indel. Primers of SAT-F and SAT-2R made a 588-bp-sized amplicon only from opium poppy. This amplicon was named "SAT2" (Table 5).

The COR region was amplified from five species (Table 5). *P. somniferum* produced two amplicons of approximately 1.4 kbp and 1.7 kbp. *P. nudicaule* and *P. rhoeas* produced amplicons of approximately 1.4 kbp, and *P. bracteatum* and *P. orientale* produced amplicons of approximately 1.7 kbp. Their COR clones were sequenced, and their sequences were registered into NCBI (Table 1). Three introns were detected at spaces between 323–324 bp, 569–570, and 749–750 in our COR clones (FJ596160 and FJ596161) from opium poppy, which were based on allele 1 of COR (AF108432). These intron lengths were 144 bp, 392 bp, and 84 bp, respectively. Two 1324-bp-sized CORs (FJ596158 and FJ596158) from opium poppy were considered to be pseudo genes, because many stop signals were found at their amino acid sequences.

Discussion

Species Identification Using ITS, 18S rRNA, rbc*L, and* trn*L*-trn*F IGS Regions*

Sequences of noncoding regions should exhibit more phylogenetically informative sites than the coding regions do, and this coincided with our study in the comparison of *trn*L-*trn*F IGS versus *rbc*L (Table 4). The evolutionary distances detected in the 14 species were highest in the comparisons of ITS sequences, followed by *trn*L-*trn*F IGS (Table 4). The sequences of ITS and *trn*L-*trn*F IGS might prove quite useful for evolutionary studies of related species. The *rbc*L and 18S rRNA sequences could be more useful for evolutionary studies at higher levels of taxonomy. Comparison of sequences with those in the NCBI database indicates that all four markers might be prominent in species identification of unknown plants, because *rbc*L and 18S rRNA allow the identification of the genus or family of an unknown plant, and ITS and *trn*L-*trn*F IGS allow species or intraspecies identification.

However, there are also some drawbacks of using universal markers for real forensic cases. First, poor coverage of plant species in the sequence databases, namely incomplete reference data, likely limits the application of the genetic approach in botanical forensics (20,21). Second, ITS sequences are not homogenized in some species (e.g., *P. rhoeas*, *C. ambigua*, and *C. incisa* in our study). These nucleotide length/sequence polymorphisms might complicate interpretation of the results. Nuclear ribosomal ITS of some plants did not work well as plant DNA barcodes because of the problems of paralogy and other factors associated with the complex concerted evolution (22). A third limitation is that ITS primer pairs can simultaneously amplify the ITS of plants as well as of endophytic or contaminating fungi (4,23) and contaminating human secretions (24). This makes it difficult to obtain nucleotide sequences via the direct sequencing. This drawback can be eliminated with plant-specific primer design (25,26). These last two phenomena have the potential to provide links between crime scenes and botanical evidences, which have heterogeneous ITS sequences or have been contaminated with geographically specific fungi or human in concern.

Opium Poppy Discrimination from Poppy Plants Using Papaver Genera-Specific Genetic Markers

This study utilized TYDC, NCS, CYP80B1, BBE, SAT, and COR genetic loci as genetic markers with the aim of making the approach feasible for drug enforcement efforts. The assessment criterion for the *Papaver* genera-specific markers was whether they produced the expected amplified product when amplified with primers designed to detect opium poppy. The TYDC, NCS, and BBE markers showed opium poppy specificity, while CYP80B1, SAT, and COR PCR products appeared in some species included in the *Papaver* genus. The presence or absence of amplified products offers a profiling mechanism to distinguish the opium poppy from different species that share similar external characteristics. Various molecular markers in chloroplast DNA, mitochondrial DNA, and nuclear DNA have tremendous utility for forensic purposes because of their species-diagnostic power (27). A robust system with a set of putatively informative indels in grass mitochondrial genome has been developed for forensic purposes (28). The opium poppy has been studied less than another narcotic plant, marijuana, which has been classified using various molecular methods such as random amplified polymorphic DNA (29), amplified fragment length polymorphism (30), inter-simple sequence repeat amplification (31), sequencing of ITS1 regions (32), and microsatellites (33,34).

DNA of rice, honeysuckle, bamboo, cypress, clover, and 9947A did not show any amplicons with six *Papaver* genera-specific markers (data not shown). Japanese hop displayed a nonspecific faint band about 500 bp in size only with the COR marker. Of 20 mg of several tissues (fresh and dried leaf, fresh root, fresh and dry stem, and seeds), fresh leaf had the largest quantity of DNA while dried stem had the least. The *Papaver* genera-specific amplicon was produced equally from every tissue (10 ng/μL of template DNA). But amplification failure occurred in the case of dried stem. Every tissue except dried stem is acceptable for opium poppy discrimination using our markers. The short amplicons (≤600 bp) such as SAT2, CYP80B1, and BBE were produced more abundantly than long amplicons (≥1.3 kbp) such as TYDC, SAT, and COR (data not shown).

Papaver radicatum, also known as *P. coreanum*, was collected from Mt. Baekdu in September 2005. It inhabits only Mt. Baekdu and is regarded as an endemic species in Korea. Another endemic species is Mae-mi-kkot, which inhabits Chonnam province central to Mt. Jiri. Having reliable genetic information concerning these endemic species would be useful in preservation efforts. The seeds of the *Papaver* family, which are numerous, have been imported for gardening or horticulture in South Korea. Using 20 mg batches of seeds or 2 to 3-cm-long seedlings, we were able to identify the species of origin using the DNA-based methods presently described. The TYDC, NCS, CYP80B1, BBE, SAT, and COR genetic markers may prove useful in forensic cases aimed at narcotics control.

Acknowledgments

This work using opium poppy was authorized by the Korea Food and Drug Administration. The authors kindly thank the members of Drug & Toxicology sector in the Western District office of the National Institute of Scientific Investigation for the offer of opium poppy specimens. This study was supported by research and development expenditures of National Institute of Scientific Investigation and Korea Evaluation Institute of Industrial Technology (B0009842-2008-02).

References

1. Ziegler HW, Beasley TH Sr, Smith DW. Simultaneous assay for six alkaloids in opium, using high-performance liquid chromatography. J Assoc Off Anal Chem 1975;58(5):888–97.

2. Tsai L-C, Hsieh H-M, Huang L-H, Wang J-C, Linacre A, Lee JC-I. Cannabis seed identification by chloroplast and nuclear DNA. Forensic Sci Int 2006;158:250–1.
3. Linacre A, Thorpe J. Detection and identification of cannabis by DNA. Forensic Sci Int 1998;91:71–6.
4. White TJ, Bruns T, Lee S, Taylor JW. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ, editors. PCR protocols: a guide to methods and applications. New York, NY: Academic Press, Inc, 1990;315–22.
5. Soltis DE, Soltis PS, Nickrent DL, Johnson LA, Hahn WJ, Hoot SB, et al. Angiosperm phylogeny inferred from 18S ribosomal DNA sequence. Ann Mo Bo Gard 1997;84(1):1–49.
6. Clegg MT. Chloroplast gene sequences and the study of plant evolution. Proc Natl Acad Sci USA 1993;90:363–7.
7. Gielly L, Taberlet P. The use of chloroplast DNA to resolve plant phylogenies: noncoding versus *rbc*L sequences. Mol Biol Evol 1994;11:769–77.
8. Taberlet P, Gielly L, Pautou G, Bouvet J. Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol Biol 1991;17:1105–9.
9. Sang T, Crawford DJ, Stuessy TF. Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). Am J Bot 1997;84:1120–36.
10. Facchini PJ, De LucaV. Phloem-specific expression of tyrosine/dopa decarboxylase genes and the biosynthesis of isoquinoline alkaloids in opium poppy. Plant Cell 1995;7:1811–21.
11. Samanani N, Liscombe DK, Facchini PJ. Molecular cloning and characterization of norcoclaurine synthase, an enzyme catalyzing the first committed step in benzylisoquinoline alkaloid biosynthesis. Plant J 2004;40:302–13.
12. Pauli HH, Kutchan TM. Molecular cloning and functional heterologous expression of two alleles encoding (S)-N-methylcoclaurine 3′-hydroxylase (CYP80B1), a new methyl jasmonate-inducible cytochrome P-450-dependent mono-oxygenase of benzylisoquinoline alkaloid biosynthesis. Plant J 1998;13:793–801.
13. Dittrich H, Kutchan TM. Molecular cloning, expression, and induction of berberine bridge enzyme, an enzyme essential to the formation of benzophenanthridine alkaloids in the response of plants to pathogenic attack. Proc Natl Acad Sci USA 1991;88:9969–73.
14. Grothe T, Lenz R, Kutchan TM. Molecular characterization of the salutaridinol 7-*O*-acetyltransferase involved in morphine biosynthesis in opium poppy *Papaver somniferum*. J Biol Chem 2001;276:30717–23.
15. Unterlinner B, Lenz R, Kutchan TM. Molecular cloning and functional expression of codeinone reductase: the penultimate enzyme in morphine biosynthesis in the opium poppy *Papaver somniferum*. Plant J 1999; 18:465–75.
16. Rollo F, Ubaldi M, Ermini L, Marota I. Otzi's last meals: DNA analysis of the intestinal content of the Neolithic glacier mummy from the Alps. Proc Natl Acad Sci USA 2002;99:12594–9.
17. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 2007;24:1596–9.
18. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25:3389–402.
19. Kumar S, Gadagkar SR. Disparity Index: a simple statistic to measure and test the homogeneity of substitution patterns between molecular sequences. Genetics 2001;158:1321–7.
20. Wesselink M, Kuiper I. Species identification of botanical trace evidence using molecular markers. Forensic Sci Int Genet 2008;1:630–2.
21. Ferri G, Alù M, Corradini B, Angot A, Beduschi G. Land plants identification in forensic botany: multigene barcoding approach. Forensic Sci Int Genet 2008;1:593–5.
22. Chase MW, Salamin N, Wilkinson M, Dunwell JM, Kesanakurthi RP, Haidar N, et al. Land plants and DNA barcodes: short-term and long-term goals. Philos Trans R Soc Lond B Biol Sci 2005;360(1462):1889–95.
23. Alvarez I, Wendel JF. Ribosomal ITS sequences and plant phylogenetic inference. Mol Phylogenet Evol 2003;29:417–34.
24. Gonzalez IL, Wu S, Li WM, Kuo BA, Sylvester JE. Human ribosomal RNA intergenic spacer sequence. Nucleic Acids Res 1992;20:5846.
25. Zhang W, Wendel JF, Clark LG. Bamboozled again! Inadvertent isolation of fungal rDNA sequences from bamboos (Poaceae: Bambusoideae). Mol Phylogenet Evol 1997;8:205–17.
26. Cullings KW, Vogler DR. A 5.8S nuclear ribosomal RNA gene sequence database: applications to ecology and evolution. Mol Ecol 1998;7:919–23.
27. Avise JC. Molecular markers, natural history and evolution, 2nd edn. Sunderland, MA: Sinauer Associates, Inc., 2004.
28. Ward J, Peakall R, Gilmore SR, Robertson J. A molecular identification system for grasses: a novel technology for forensic botany. Forensic Sci Int 2005;152:121–31.
29. Gillan R, Cole MD, Linacre A, Thorpe JW, Watson ND. Comparison of Cannabis sativa by random amplification of polymorphic DNA (RAPD) and HPLC of cannabinoid: a preliminary study. Sci Justice 1995;35: 169–77.
30. Miller CoyleH, Shutler G, Abrams S, Hanniman J, Neylon S, Ladd C, et al. A simple DNA extraction method for marijuana samples used in amplified fragment length polymorphism (AFLP) analysis. J Forensic Sci 2003;48(2):343–7.
31. Kojoma M, Iida O, Makino Y, Sekita S, Satake M. DNA fingerprinting of *Cannabis sativa* using inter-simple sequence repeat (ISSR) amplification. Planta Med 2002;68:60–3.
32. Gigliano GS. Preliminary data on the usefulness of internal transcribed spacer 1 (ITS1) sequence in *Cannabis sativa* L identification. J Forensic Sci 1999;44:457–75.
33. Hsieh H-J, Hou R-J, Tsai L-C, Wei C-S, Liu S-W, Huang L-H, et al. A highly polymorphic STR locus in *Cannabis sativa*. Forensic Sci Int 2003;131(1):53–8.
34. Howard C, Gilmore S, Robertson J, Peakall R. Developmental validation of a *Cannabis sativa* STR multiplex system for forensic analysis. J Forensic Sci 2008;53(5):1061–7.

Additional information and reprint requests:
Eun-jung Lee, Ph.D.
DNA Analysis Sector
Western District Office of National Institute of Scientific Investigation
111 Daeduk-Ri, Seosam-Myun
Jangsung-Gun, Chonnam
515-822, South Korea
E-mail: ejlee36@nisi.go.kr

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01441.x
Available online at: interscience.wiley.com

PAPER

CRIMINALISTICS

Zlatko Mehmedic,[1] *M.Sc.Pharm.; Suman Chandra,*[1] *Ph.D.; Desmond Slade,*[1] *Ph.D.; Heather Denham,*[1] *B.A.; Susan Foster,*[1] *B.A.; Amit S. Patel,*[2,3] *Ph.D.; Samir A. Ross,*[1,4] *Ph.D.; Ikhlas A. Khan,*[1,4] *Ph.D.; and Mahmoud A. ElSohly,*[1,5] *Ph.D.*

Potency Trends of Δ^9-THC and Other Cannabinoids in Confiscated Cannabis Preparations from 1993 to 2008*

ABSTRACT: The University of Mississippi has a contract with the National Institute on Drug Abuse (NIDA) to carry out a variety of research activities dealing with cannabis, including the Potency Monitoring (PM) program, which provides analytical potency data on cannabis preparations confiscated in the United States. This report provides data on 46,211 samples seized and analyzed by gas chromatography-flame ionization detection (GC-FID) during 1993–2008. The data showed an upward trend in the mean Δ^9-tetrahydrocannabinol (Δ^9-THC) content of all confiscated cannabis preparations, which increased from 3.4% in 1993 to 8.8% in 2008. Hashish potencies did not increase consistently during this period; however, the mean yearly potency varied from 2.5–9.2% (1993–2003) to 12.0–29.3% (2004–2008). Hash oil potencies also varied considerably during this period (16.8 ± 16.3%). The increase in cannabis preparation potency is mainly due to the increase in the potency of nondomestic versus domestic samples.

KEYWORDS: cannabichromene (CBC), cannabidiol (CBD), cannabigerol (CBG), cannabinoids, cannabinol (CBN), cannabis, criminalistics, forensic science, gas chromatography-flame ionization detection (GC-FID), marijuana, potency, tetrahydrocannabivarin (THCV), Δ^9-tetrahydrocannabinol (Δ^9-THC)

Marijuana, the crude drug derived from *Cannabis sativa* L. pistillate inflorescence, is the most widely cultivated and consumed illicit drug in the world despite being under international control for eight decades (1,2). The reason for this is mainly attributed to two factors; namely, relaxation of cannabis law enforcement relative to other illicit drugs and the enormous extent of cannabis production and consumption. Furthermore, cannabis is cultivated both indoors and outdoors, often on a small scale, facilitating inconspicuous trading. Hashish (hash) and hash oil are two preparations designed to minimize the volume of the drug, thereby minimizing confiscation.

The Δ^9-tetrahydrocannabinol (Δ^9-THC) potency (concentration or content) of cannabis depends on soil and climate conditions, variety (phenotype), and cultivation techniques, with different parts of the plant having varying concentrations of the drug (3–6). The total number of identified cannabis constituents has increased from 489 in 2005 (7) to 537 in 2009, while the number of cannabinoids has increased from 70 to 109 (8–13). The main psychoactive ingredient in cannabis is Δ^9-THC (14,15); however, other cannabinoids have also demonstrated pharmacological activities, e.g., the nonpsychotropic cannabinoid cannabidiol (CBD) displays antipsychotic, antihyperalgesic, anticonvulsant, neuroprotective, and antiemetic properties (16–18).

The complex political, medical, cultural, and socioeconomic issues associated with cannabis necessitates not only public and governmental scrutiny, but especially scientific inquiry (1,2,19–24). The National Institute on Drug Abuse (NIDA) Potency Monitoring (PM) program at the National Center for Natural Products Research, University of Mississippi, provides analytical potency data on cannabis preparations seized in the United States, including both domestic and nondomestic material (25–28). A survey of the literature reporting similar programs in other countries revealed a number of comprehensive studies, e.g., England (2004–2005) (29), Brazil (2006–2007) (30), Netherlands (1999–2007) (31–34), Italy (1997–2004) (35), New Zealand (1976–1996) (36), and Australia (37), as well as a number of general reviews pertaining to cannabis potency trends (1,2,21,22,32,38,39).

This report covers 46,211 cannabis preparations confiscated and analyzed by gas chromatography-flame ionization detection (GC-FID) in the United States during 1993–2008, following on previous reports covering 1972–1997 (36,297 samples) (25–28). The total number of samples received during this period (1993–2008) was 47,583 as of 30 March 2009. The number of samples analyzed was 46,211, with 1,372 samples not analyzed for a variety of reasons, including insufficient material, wet material, and material containing only seeds and stems. Statistical analysis on the mean yearly Δ^9-THC concentration is included to establish the potency trend over time. Data on hashish, hash oil, and the potencies of

[1]National Center for Natural Products Research, School of Pharmacy, University of Mississippi, University, MS 38677.
[2]Department of Pharmacy Administration, School of Pharmacy, University of Mississippi, University, MS 38677.
[3]Current address: Medical Marketing Economics, LLC, PO Box 2309, Oxford, MS 38655.
[4]Department of Pharmacognosy, School of Pharmacy, University of Mississippi, University, MS 38677.
[5]Department of Pharmaceutics, School of Pharmacy, University of Mississippi, University, MS 38677.
*This project was supported by the National Institute on Drug Abuse (contract number N01DA-5-7746).

Received 15 May 2009; and in revised form 14 July 2009; accepted 31 July 2009.

cannabichromene (CBC), cannabidiol (CBD), cannabinol (CBN), cannabigerol (CBG), and tetrahydrocannabivarin (THCV) are also presented.

Materials and Methods

Sample Acquisition

All samples analyzed in this investigation were confiscated during 1993 through 2008 by United States Federal and State law enforcement agencies.

Sample Identification

Sample classification is based on physical characteristics according to the following guidelines:

Cannabis Samples—All samples were received as raw plant material. These samples were further categorized as follows:

- *Marijuana* (known as herbal cannabis in Europe): usually found in four forms: (i) loose material - loose cannabis plant material with leaves, stems, and seeds; (ii) leaves - cannabis plant material consisting primarily of leaves; (iii) kilo bricks - compressed cannabis with leaves, stems, and seeds (typical Mexican packaging); and (iv) buds - flowering tops of female plants with seeds.
- *Sinsemilla*: flowering tops of unfertilized female plants with no seeds (subdivided as for marijuana with most samples being classified as buds).
- *Thai sticks*: leafy material tied around a small stem (typical Thailand packaging).
- *Ditchweed*: fiber type wild cannabis found in the Midwestern region of the United States (subdivided as for marijuana).

Hashish Samples—Hashish (known as cannabis resin in Europe) is composed of the resinous parts of the flowering tops of cannabis, mixed with some plant particles and shaped into a variety of forms, e.g., balls, sticks, or slabs. It is generally very hard with a dark green or brownish color.

Hash Oil Samples—Hash oil is a liquid or semi-solid concentrated extract of cannabis plant material. Depending on the process used to prepare hash oil, it is usually dark green, amber, or brownish.

Sample Storage

All samples are stored in a vault at controlled room temperature (17 ± 4°C).

Domestically Cultivated Cannabis

Cannabis preparations that have been verified as being produced from plants grown in the United States are classified as domestic samples, whereas all other samples are classified as nondomestic.

Sample Preparation

Cannabis—The samples were manicured in a 14 mesh metal sieve to remove seeds and stems. Duplicate samples (2 × 0.1 g) were extracted with internal standard solution (ISTD) [3 mL, 4–androstene-3,17-dione (100 mg) (Sigma Aldrich, St. Louis, MO) in chloroform/methanol (100 mL, 1:9, v/v), 1 mg/mL] at room temperature for 1 h. The extracts were transferred to GC vials via filtration through sterile cotton plugs, followed by capping of the vials (25).

Hashish—Samples were powdered using a mortar and pestle or an electric blender. Duplicate samples (2 × 0.1 g) were extracted following the procedure outlined for cannabis samples (*vide supra*).

Hash Oil—Duplicate samples (2 × 0.1 g) were extracted with ISTD [4 mL, 4-androstene-3,17-dione (50 mg) in absolute ethanol (50 mL), 1 mg/mL] as follows: maceration at room temperature for 2–4 h, sonication for 5 min, addition of absolute ethanol (20 mL), and sonication for 5 min. The extracts were transferred to GC vials as described earlier.

Chromatographic Analysis

GC analyses were performed using Varian CP-3380 gas chromatographs, equipped with Varian CP-8400 automatic liquid samplers, capillary injectors, dual flame ionization detectors, and DB-1MS columns (15 m × 0.25 mm × 0.25 μm) (J&W Scientific, Folsom, CA). Data were recorded using a Dell Optiplex GX1 computer and Varian Star workstation software (version 6.1). Helium was used as carrier and detector makeup gas with an upstream indicating moisture trap and a downstream indicating oxygen trap. Hydrogen and compressed air were used as the combustion gases. The following instrument parameters were employed: air, 30 psi (300 mL/min); hydrogen, 30 psi (30 mL/min); column head pressure, 14 psi (1.0 mL/min); split flow rate, 100 mL/min; split ratio, 50:1; septum purge flow rate: 5 mL/min; makeup gas pressure, 20 psi (30 mL/min); injector temperature, 240°C; detector temperature, 270°C; oven program, 170°C (hold 1 min) to 250°C at 10°C/min (hold 3 min); run time, 12 min; injection volume, 1 μL. The instruments are daily maintained and calibrated to ensure a Δ^9-THC/internal standard response factor ratio of one.

Calculation of Concentrations

The concentration of a specific cannabinoid is calculated as follows:

$$cannabinoid\% = \frac{GC[area](cannabinoid)}{GC[area](ISTD)} \times \frac{amount(ISTD)}{amount(sample)} \times 100$$

Statistical Analysis

The mean and standard deviation (SD) of the sample concentrations were calculated for the combined data set, by year and sample type, and for domestic and nondomestic samples. Normal and outlier cannabis samples were determined based on the mean and SD of the Δ^9-THC concentration for each year and sample type (40). Normal samples are defined as samples with potencies in the range: mean ± 2.5 × SD. Outlier samples are defined as samples with potencies that fall outside this range. The precision of the mean was determined through 95% confidence intervals (CIs). The CI was calculated using the Excel function TINV(probability, degrees of freedom), which returns the inverse or t-value of the Student's t-distribution as a function of the probability associated with the two-tailed Student's t-distribution and the degrees of freedom [number of samples (n) – 1]. The CI range is subsequently calculated as the mean ± the product of the TINV value and the standard error of the mean (SEM), i.e., the SD divided by the square root of the number of samples, thus mean ± SEM × TINV

[$SEM = SD/\sqrt{n}$, TINV = TINV(0.05, $n - 1$)]. A 95% CI is a range of values that contains the true mean of the population with 95% certainty. The Pearson product-moment correlation coefficient (r) was calculated using the Excel PEARSON function, and the standard error for the predicted mean values for each year in the regression was calculated using the Excel STEYX function.

Results and Discussion

During the past 16 years (1993–2008), 46,211 samples of cannabis preparations confiscated in the United States, representing *c.* 8,321 tons, were analyzed at the University of Mississippi PM laboratory (Table 1). The PM program has analyzed 67,227 samples to date since 1968 (25–28). Samples classification is performed by the submitting agency and verified by the PM laboratory. Prior to 1995, there was no classification in the database for ditchweed; therefore, all ditchweed samples were classified as marijuana. However, interest in monitoring ditchweed samples and its effect on the overall potency of confiscated marijuana necessitated this category on the sample report form since 1995. The data presented in this report on ditchweed samples prior to 1995 were generated by retrospective review of the PM data. Marijuana samples with Δ^9-THC <1% and CBD > Δ^9-THC were classified as ditchweed. Cannabis, i.e., marijuana, sinsemilla, Thai sticks, and ditchweed, represents the overwhelming majority of the samples confiscated in the United States (98.7%), while the hashish and hash oil combined contribution is <1.5% (Table 1). Marijuana typically represents at least 50% of the samples. Sinsemilla samples gradually increased from 2002, with a concurrent decrease in the number of marijuana samples.

The yearly arithmetic mean Δ^9-THC concentration for the different types of cannabis samples shows large variation within categories and over time, with only the ditchweed samples being relatively constant (Table 2). Hashish and hash oil sample potencies

TABLE 1—*Number of samples (*n*) analyzed by type and year.*

	All	Marijuana*		Sinsemilla*		Thai sticks*		Ditchweed*		Hashish†		Hash oil†	
Year	*n*	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
1993	3412	3033	88.9	123	3.6	0	0.0	200	5.9	39	1.1	17	0.5
1994	3327	3032	91.1	104	3.1	0	0.0	148	4.4	29	0.9	14	0.4
1995	4791	4430	92.5	164	3.4	2	0.04	163	3.4	19	0.4	13	0.3
1996	2455	2148	87.5	169	6.9	0	0.0	118	4.8	12	0.5	8	0.3
1997	2495	2273	91.1	121	4.8	0	0.0	60	2.4	31	1.2	10	0.4
1998	2283	2075	90.9	101	4.4	0	0.0	87	3.8	15	0.7	5	0.2
1999	2692	2450	91.0	136	5.1	0	0.0	72	2.7	23	0.9	11	0.4
2000	3148	2928	93.0	113	3.6	0	0.0	73	2.3	27	0.9	7	0.2
2001	2716	2398	88.3	235	8.7	0	0.0	63	2.3	13	0.5	7	0.3
2002	2413	1789	74.1	528	21.9	0	0.0	75	3.1	16	0.7	5	0.2
2003	2517	1893	75.2	538	21.4	0	0.0	66	2.6	16	0.6	4	0.2
2004	2637	1815	68.8	731	27.7	0	0.0	62	2.4	25	0.9	4	0.2
2005	3004	1964	65.4	931	31.0	0	0.0	56	1.9	47	1.6	6	0.2
2006	2890	1770	61.2	1032	35.7	0	0.0	53	1.8	32	1.1	3	0.1
2007	3097	1635	52.8	1327	42.8	0	0.0	47	1.5	70	2.3	18	0.6
2008	2334	1151	49.3	1093	46.8	0	0.0	28	1.2	50	2.1	12	0.5
1993–2008	46,211	36,784	79.6	7446	16.1	2	0.0	1371	3.0	464	1.0	144	0.3

*Total cannabis: 45,603 samples (98.7%).
†Total hashish + hash oil: 608 samples (1.3%).

TABLE 2—*Mean and SD Δ^9-THC concentration by type of sample and year.*

	All		Marijuana		Sinsemilla		Thai sticks		Ditchweed		Hashish		Hash oil	
Year	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1993	3.4	2.9	3.4	2.4	5.8	3.8	0.0	0.0	0.4	0.3	6.6	6.7	16.5	11.7
1994	3.5	2.5	3.5	2.1	7.5	4.8	0.0	0.0	0.4	0.3	4.6	3.6	11.6	7.9
1995	3.8	2.3	3.7	1.8	7.5	4.4	4.5	0.8	0.4	0.4	3.6	3.7	13.2	8.9
1996	4.1	3.0	3.9	2.2	9.2	4.7	0.0	0.0	0.4	0.3	2.5	1.4	12.8	9.5
1997	4.6	3.7	4.3	2.7	11.6	5.9	0.0	0.0	0.5	0.3	8.9	9.3	18.2	9.0
1998	4.5	3.6	4.2	2.9	12.3	5.2	0.0	0.0	0.4	0.3	5.9	5.2	15.8	9.9
1999	4.6	4.0	4.2	3.2	13.4	4.7	0.0	0.0	0.4	0.3	4.9	4.2	16.2	10.7
2000	4.9	4.0	4.7	3.4	12.8	4.4	0.0	0.0	0.4	0.3	4.2	4.2	28.6	11.6
2001	5.4	4.1	5.0	3.5	9.6	5.4	0.0	0.0	0.4	0.3	8.5	5.9	19.4	8.1
2002	6.4	5.1	5.1	3.4	11.4	5.7	0.0	0.0	0.4	0.3	9.1	8.5	22.5	28.3
2003	6.3	4.8	5.0	3.1	11.6	5.7	0.0	0.0	0.3	0.3	9.2	7.6	15.5	6.9
2004	7.2	5.8	5.4	3.6	11.9	6.0	0.0	0.0	0.4	0.3	18.9	15.1	31.3	34.6
2005	7.2	5.3	5.2	3.2	11.6	5.7	0.0	0.0	0.4	0.3	12.0	10.3	6.4	2.8
2006	7.8	6.5	5.6	4.0	11.2	6.5	0.0	0.0	0.3	0.2	29.3	19.7	18.7	26.1
2007	8.8	7.4	6.1	3.7	11.1	6.6	0.0	0.0	0.4	0.3	27.7	18.4	24.9	29.6
2008	8.8	6.9	5.8	3.9	11.5	6.2	0.0	0.0	0.4	0.3	23.1	19.6	6.5	9.7
1993–2008	5.6	5.0	4.5	3.1	11.1	6.1	4.5	0.8	0.4	0.3	14.1	15.7	16.8	16.3
95% CI range*	5.53–5.62		4.46–4.53		11.01–11.28		0.00–11.69		0.37–0.40		12.69–15.56		14.07–19.45	

SD, Standard deviation.
*95% CI range: range of values that contains the true mean with 95% certainty.

showed the most variability over the 16-year period. The mean and SD for these categories were 14.1% ± 15.7% and 16.8% ± 16.3%, respectively. The marijuana Δ^9-THC concentration appeared to gradually increase from 1993 to 2008, with a Pearson product-moment correlation coefficient (r) of 0.982 and a standard error for the predicted mean values of 0.17 (Fig. 1). The mean Δ^9-THC concentration for sinsemilla fluctuated considerably, ranging from a minimum in 1993 (5.8% ± 3.8%) to a maximum in 1999 (13.4% ± 4.7%) (Table 2, Fig. 1). Other than the expected finding that the yearly mean potencies of sinsemilla samples were much higher than that for marijuana samples, there did not appear to be any meaningful trend in the mean potency of the sinsemilla samples. The mean Δ^9-THC concentration of sinsemilla samples between 1993 and 2000 increased from 5.8% to 12.8% (121.8% increase), dropping slightly in 2001 (9.6%), and stabilizing between 2002 and 2008 (11.5% ± 0.3%) (Fig. 1).

The change in cannabis potency over the past 40 years has been the subject of much debate and controversy. This report investigates the influence of outlier samples on the overall mean concentration of Δ^9-THC for the time period studied in an attempt to clarify this issue. Normal and outlier cannabis preparations are samples with Δ^9-THC concentrations that fall within and outside the range mean ± 2.5 × SD, respectively.

The outlier samples for marijuana and sinsemilla represent 2.4% and 0.5%, respectively, of the total samples for each type (Table 3). The distribution of Δ^9-THC concentrations is positively skewed,

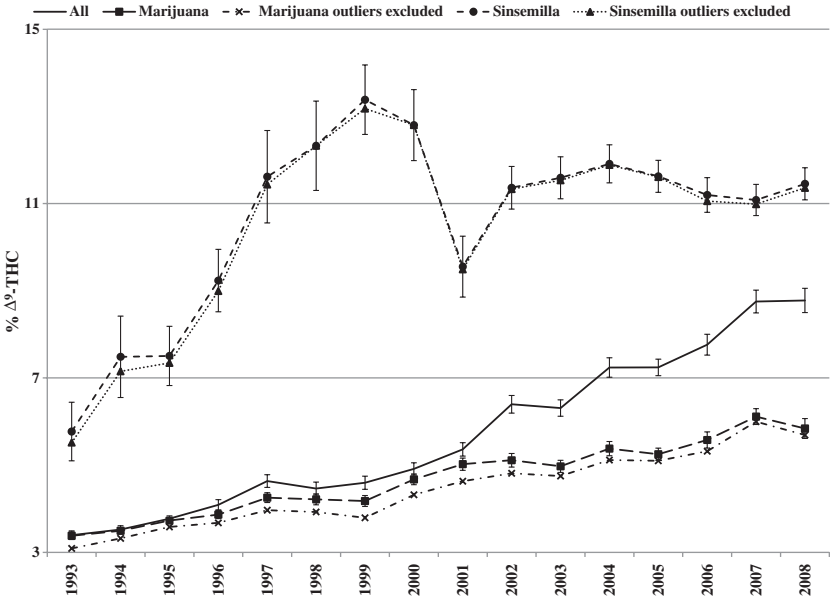


FIG. 1—*Mean Δ^9-THC concentration with 95% confidence intervals for all samples, marijuana and sinsemilla samples, and marijuana and sinsemilla samples with outliers excluded.*

TABLE 3—*Mean and SD Δ^9-THC concentration for marijuana and sinsemilla samples with outliers* excluded.*

	Marijuana					Sinsemilla				
	Outliers	All samples		Outliers excluded		Outliers	All samples		Outliers excluded	
Year	%	Mean	SD	Mean	SD	%	Mean	SD	Mean	SD
1993	2.9	3.4	2.4	3.1	1.7	2.4	5.8	3.8	5.5	3.4
1994	2.3	3.5	2.1	3.3	1.7	1.9	7.5	4.8	7.2	4.2
1995	2.0	3.7	1.8	3.6	1.5	1.2	7.5	4.4	7.3	4.2
1996	2.3	3.9	2.2	3.7	1.8	1.8	9.2	4.7	9.0	4.4
1997	3.1	4.3	2.7	4.0	2.2	0.8	11.6	5.9	11.4	5.6
1998	2.7	4.2	2.9	3.9	2.3	0.0	12.3	5.2	12.3	5.2
1999	3.5	4.2	3.2	3.8	2.4	1.5	13.4	4.7	13.2	4.4
2000	3.2	4.7	3.4	4.3	2.8	0.0	12.8	4.4	12.8	4.4
2001	3.4	5.0	3.5	4.6	2.8	0.4	9.6	5.4	9.5	5.4
2002	2.5	5.1	3.4	4.8	2.8	0.2	11.4	5.7	11.3	5.7
2003	2.1	5.0	3.1	4.8	2.7	0.4	11.6	5.7	11.5	5.6
2004	2.1	5.4	3.6	5.1	3.1	0.1	11.9	6.0	11.9	6.0
2005	1.5	5.2	3.2	5.1	3.0	0.1	11.6	5.7	11.6	5.7
2006	2.0	5.6	4.0	5.3	3.5	0.8	11.2	6.5	11.1	6.3
2007	0.9	6.1	3.7	6.0	3.5	0.5	11.1	6.6	11.0	6.5
2008	1.1	5.8	3.9	5.7	3.7	0.5	11.5	6.2	11.4	6.1
1993–2008	2.4	4.5	3.1	4.2	2.7	0.5	11.1	6.1	11.1	6.0
95% CI range†	–	4.46–4.53		4.22–4.27		–	11.01–11.28		10.92–11.20	

SD, Standard deviation.
*Mean – 2.5 × SD > Outlier > Mean + 2.5 × SD.
†95% CI range: range of values that contains the true mean with 95% certainty.

i.e., all outliers are samples with potencies higher than the mean potency. It is therefore important that the potential effect of the outliers is examined to determine whether the apparent trend of increasing potency is real or simply a statistical artifact. A comparison of the mean potency of marijuana and sinsemilla samples calculated for all samples versus for samples with outliers excluded indicates that the mean Δ^9-THC concentration decreases for each year when the outliers are excluded (Table 3, Fig. 1). However, the general pattern of increasing potency of marijuana samples since 1993 appears to exist even when outliers are excluded. The Pearson product-moment correlation coefficient (*r*) and standard error for the predicted mean values after exclusion of marijuana sample outliers were 0.981 and 0.18, respectively. Because of the greater variability found in the potency of sinsemilla samples, fewer cases were excluded as outliers and thus there was little effect on the mean potency for each of the years reported (Table 3, Fig. 1). The mean Δ^9-THC concentration for marijuana and sinsemilla samples decreased by 0.24% and 0.08%, respectively, after exclusion of the outliers.

Further evidence that the mean Δ^9-THC concentration for marijuana may be increasing is inferred by the analysis of the percentage of samples each year with Δ^9-THC concentration more than 3%, 5%, and 9%. Marijuana samples with Δ^9-THC >9% increased from 3.23% (1993) to a maximum 21.47% (2007). Conversely, the number of marijuana sample containing Δ^9-THC <3% decreased between 1993 and 2007, with a slight increase in 2008 (Fig. 2). The trend for sinsemilla samples with Δ^9-THC >9% followed a similar pattern to the overall trend for the yearly mean potencies

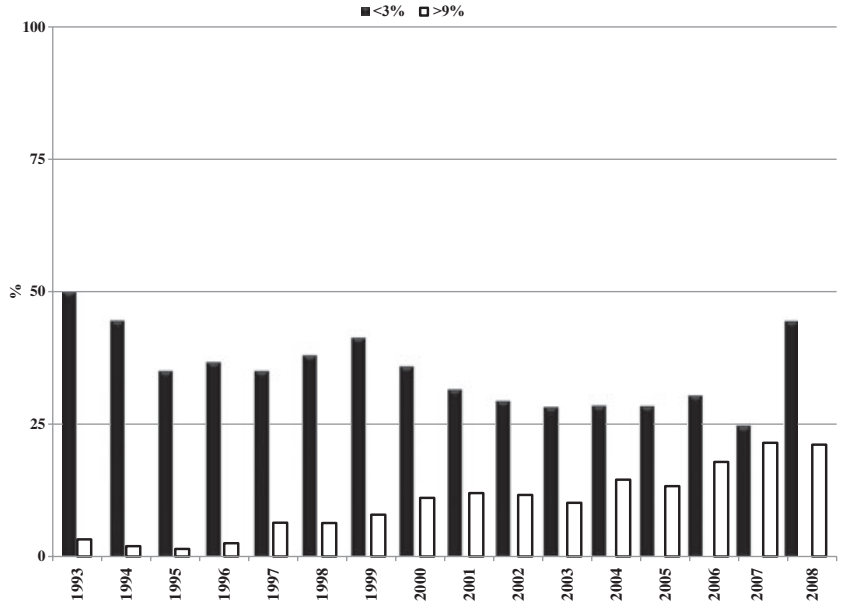


FIG. 2—*Prevalence of low (<3%) and high (>9%) potency marijuana samples.*

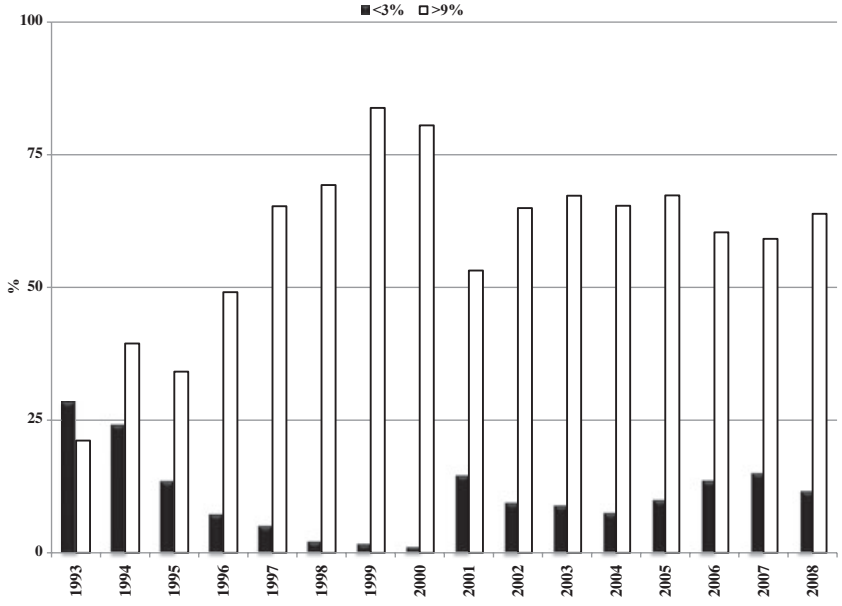


FIG. 3—*Prevalence of low (<3%) and high (>9%) potency sinsemilla samples.*

(Figs 1 and 3). Considering the large number of cannabis samples analyzed each year, it is doubtful that these observations are statistical artifacts.

The overall number of samples, mean, SD, maximum and minimum concentrations of Δ^9-THC for the different types of samples categorized by origin, i.e., domestic or nondomestic, indicates that ditchweed is mainly a domestic product, whereas Thai sticks, hashish, and hash oil are nondomestic products (Table 4). Marijuana and sinsemilla samples represent more than 95% of all seizures. It is important to mention that samples are classified as being of domestic origin only if the seizure is made from a growing operation (indoor or outdoor) within the United States. All other samples are classified as being nondomestic, although they could possibly have been produced in the United States prior to seizure. It is also important to note that all nondomestic sample seizures made by the DEA are of final products produced from mature plant material. In contrast, the domestic samples provided by the state eradication programs are seized at different stages of plant maturity. Overall, the number of samples of known domestic origin represents approximately one-third of all samples confiscated. The number of nondomestic seizures was consistently higher when compared to that of domestic seizures (Fig. 4). The mean Δ^9-THC concentration for nondomestic cannabis samples showed a gradual increase, while domestic samples had little fluctuation (Fig. 5).

The mean concentration of the minor cannabinoids CBC, CBD, CBN, CBG, and THCV were also monitored (Table 5). CBD is the major cannabinoid found in ditchweed and is present in elevated amounts in intermediate type cannabis (moderate levels of both Δ^9-THC and CBD) used to make hashish. The cannabinoid content of hashish and hash oil samples shows that, while hashish

TABLE 4—*Number of samples (*n*), mean, SD, maximum and minimum Δ^9-THC concentration by origin and type of sample.*

Origin	Type	*n*	Mean	SD	Maximum	Minimum
Domestic	Marijuana	10,308	3.0	2.8	24.7	<0.01
	Sinsemilla	3067	7.9	5.5	33.1	0.1
	Thai sticks	0	–	–	–	–
	Ditchweed	1257	0.4	0.3	2.4	<0.01
	Hashish	3	34.0	25.4	52.9	5.1
	Hash oil	2	0.2	0.01	0.23	0.21
	1993–2008	14,637	3.8	4.1	52.9	<0.01
Nondomestic	Marijuana	26,476	5.1	3.0	37.2	<0.01
	Sinsemilla	4379	13.4	5.4	32.3	0.5
	Thai sticks	2	4.5	0.8	5.1	4.0
	Ditchweed	114	0.4	0.3	1.2	0.1
	Hashish	461	14.0	15.6	66.3	<0.01
	Hash oil	142	17.0	16.3	81.7	<0.01
	1993–2008	31,574	6.4	5.1	81.7	<0.01
All Samples	Marijuana	36,784	4.5	3.1	37.2	<0.01
	Sinsemilla	7446	11.1	6.1	33.1	0.1
	Thai sticks	2	4.5	0.8	5.1	4.0
	Ditchweed	1371	0.4	0.3	2.4	<0.01
	Hashish	464	14.1	15.7	66.3	<0.01
	Hash oil	144	16.8	16.3	81.7	<0.01
	1993–2008	46,211	5.6	5.0	81.7	<0.01

SD, Standard deviation.

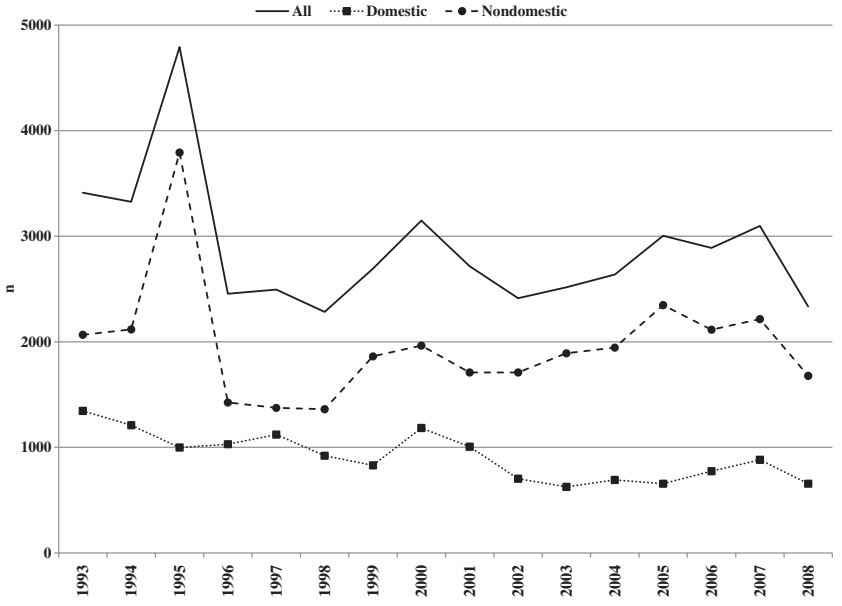


FIG. 4—*Number (*n*) of domestic and nondomestic samples.*

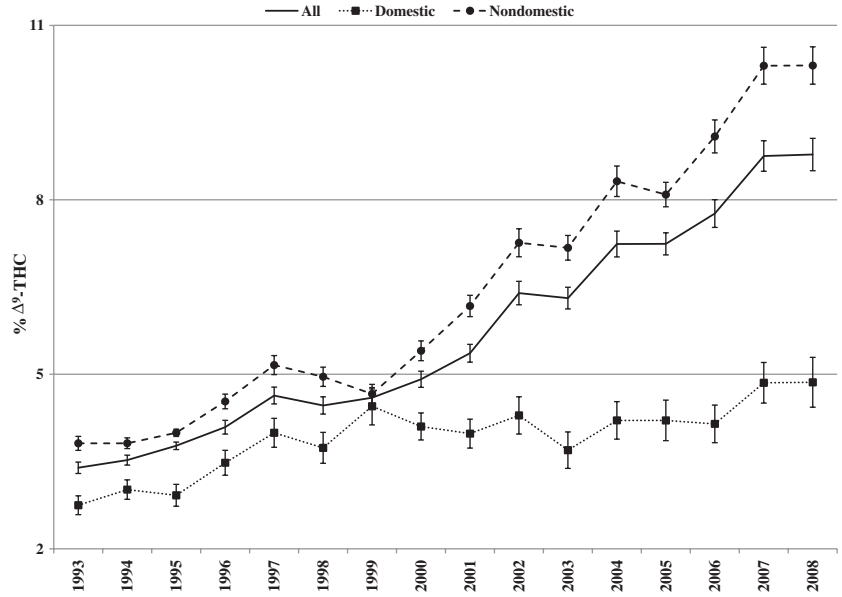


FIG. 5—*Δ^9-THC concentration of domestic and nondomestic samples with 95% confidence intervals.*

TABLE 5—*Mean concentration of minor cannabinoids by type and year.*

	All						Marijuana						Sinsemilla					
Year	THC	CBC	CBD	CBN	CBG	THCV	THC	CBC	CBD	CBN	CBG	THCV	THC	CBC	CBD	CBN	CBG	THCV
1993	3.4	0.2	0.3	0.3	0.0	0.0	3.4	0.2	0.2	0.3	0.0	0.0	5.8	0.2	0.2	0.0	0.1	0.0
1994	3.5	0.2	0.4	0.2	0.1	0.1	3.5	0.2	0.3	0.2	0.1	0.1	7.5	0.2	0.5	0.1	0.3	0.1
1995	3.8	0.2	0.3	0.3	0.1	0.0	3.7	0.2	0.3	0.3	0.1	0.0	7.5	0.3	0.3	0.1	0.3	0.1
1996	4.1	0.2	0.4	0.3	0.2	0.1	3.9	0.2	0.3	0.2	0.1	0.1	9.2	0.3	0.5	0.1	0.4	0.1
1997	4.6	0.3	0.4	0.2	0.2	0.1	4.3	0.3	0.4	0.2	0.2	0.1	11.6	0.3	0.4	0.1	0.5	0.1
1998	4.5	0.2	0.4	0.3	0.2	0.1	4.2	0.2	0.3	0.2	0.1	0.1	12.3	0.4	0.4	0.2	0.5	0.1
1999	4.6	0.2	0.4	0.4	0.2	0.0	4.2	0.2	0.4	0.4	0.2	0.0	13.4	0.3	0.3	0.2	0.5	0.1
2000	4.9	0.2	0.5	0.4	0.2	0.1	4.7	0.2	0.4	0.4	0.2	0.1	12.8	0.2	0.3	0.2	0.4	0.1
2001	5.4	0.2	0.5	0.3	0.3	0.1	5.0	0.2	0.5	0.3	0.2	0.1	9.6	0.2	0.3	0.2	0.4	0.1
2002	6.4	0.2	0.4	0.2	0.2	0.1	5.1	0.2	0.5	0.2	0.2	0.1	11.4	0.3	0.2	0.2	0.3	0.1
2003	6.3	0.2	0.5	0.2	0.3	0.1	5.0	0.2	0.5	0.3	0.3	0.1	11.6	0.3	0.3	0.2	0.4	0.1
2004	7.2	0.3	0.5	0.3	0.3	0.1	5.4	0.2	0.5	0.3	0.3	0.1	11.9	0.3	0.2	0.2	0.5	0.1
2005	7.2	0.3	0.5	0.3	0.4	0.1	5.2	0.3	0.5	0.4	0.3	0.1	11.6	0.3	0.2	0.2	0.4	0.1
2006	7.8	0.2	0.4	0.3	0.3	0.1	5.6	0.2	0.5	0.3	0.3	0.1	11.2	0.3	0.2	0.2	0.4	0.1
2007	8.8	0.3	0.4	0.3	0.4	0.1	6.1	0.2	0.5	0.3	0.3	0.1	11.1	0.3	0.3	0.2	0.4	0.1
2008	8.8	0.3	0.4	0.3	0.4	0.1	5.8	0.2	0.4	0.3	0.3	0.1	11.5	0.3	0.2	0.2	0.4	0.1
1993–2008	5.6	0.2	0.4	0.3	0.2	0.1	4.5	0.2	0.4	0.3	0.2	0.1	11.1	0.3	0.2	0.2	0.4	0.1
SD	5.0	0.3	0.9	0.5	0.3	0.1	3.1	0.2	0.7	0.4	0.3	0.1	6.1	0.4	0.9	0.3	0.4	0.1

	Ditchweed						Hashish						Hash oil					
Year	THC	CBC	CBD	CBN	CBG	THCV	THC	CBC	CBD	CBN	CBG	THCV	THC	CBC	CBD	CBN	CBG	THCV
1993	0.4	0.1	1.7	0.0	0.0	0.0	6.6	0.7	3.8	2.3	0.5	0.3	16.5	0.7	0.1	7.7	0.3	0.5
1994	0.4	0.1	2.0	0.0	0.0	0.0	4.6	0.5	3.5	1.7	0.5	0.2	11.6	0.6	0.2	3.1	0.4	0.5
1995	0.4	0.1	1.6	0.0	0.1	0.0	3.6	0.5	3.3	1.7	0.3	0.1	13.2	1.0	0.7	4.2	0.5	0.3
1996	0.4	0.1	2.1	0.0	0.1	0.0	2.5	0.7	4.5	2.4	0.3	0.1	12.8	1.1	1.3	4.0	0.5	0.5
1997	0.5	0.1	1.9	0.0	0.0	0.0	8.9	0.7	4.0	2.1	0.5	0.3	18.2	1.0	0.3	3.5	0.3	0.6
1998	0.4	0.2	2.0	0.0	0.0	0.0	5.9	0.8	1.7	2.0	0.3	0.2	15.8	0.8	0.2	3.6	0.2	0.5
1999	0.4	0.1	1.8	0.1	0.1	0.0	4.9	0.6	1.8	2.1	0.5	0.3	16.2	1.3	0.4	4.8	0.3	0.4
2000	0.4	0.1	2.0	0.0	0.0	0.0	4.2	0.6	4.9	2.3	0.4	0.1	28.6	1.6	0.5	1.7	0.9	0.7
2001	0.4	0.1	1.8	0.0	0.1	0.0	8.5	0.6	2.7	1.5	0.6	0.3	19.4	1.2	1.3	4.4	0.9	0.6
2002	0.4	0.1	1.5	0.0	0.0	0.0	9.1	0.6	2.5	1.4	0.4	0.2	22.5	0.5	0.3	1.7	1.2	0.3
2003	0.3	0.1	1.8	0.1	0.1	0.0	9.2	0.7	3.9	1.8	0.4	0.2	15.5	0.8	0.2	1.3	0.3	0.4
2004	0.4	0.1	1.5	0.1	0.1	0.0	18.9	0.7	0.8	1.4	0.7	0.2	31.3	1.1	1.1	2.2	1.2	0.4
2005	0.4	0.1	1.9	0.1	0.1	0.0	12.0	0.9	1.7	1.9	0.4	0.2	6.4	0.2	0.3	1.1	0.2	0.2
2006	0.3	0.1	2.4	0.2	0.1	0.0	29.3	0.7	1.6	1.3	0.8	0.2	18.7	0.4	0.1	0.6	0.4	0.1
2007	0.4	0.1	2.0	0.1	0.1	0.0	27.7	0.8	1.2	1.8	1.0	0.3	24.9	0.9	0.6	1.5	0.7	0.3
2008	0.4	0.2	1.9	0.0	0.1	0.0	23.1	0.9	2.1	2.1	0.9	0.4	6.5	0.3	0.2	0.8	0.2	0.1
1993–2008	0.4	0.1	1.8	0.0	0.0	0.0	14.1	0.7	2.5	1.9	0.6	0.3	16.8	0.9	0.5	3.3	0.5	0.4
SD	0.3	0.1	1.5	0.2	0.1	0.0	15.7	0.7	2.9	1.4	0.6	0.3	16.3	0.9	0.8	3.8	0.7	0.4

CBC, cannabichromene; CBD, cannabidiol; CBG, cannabigerol; CBN, cannabinol; Δ^9-THC, Δ^9-tetrahydrocannabinol; THCV, tetrahydrocannabivarin.

is prepared from intermediate type cannabis, hash oil is prepared from drug-type cannabis (high Δ^9-THC and low CBD levels) (3–6,16). CBC and CBN are usually higher in hashish and hash oil samples compared to cannabis samples. The CBN concentration relative to Δ^9-THC reflects the age of the samples (41). CBG content is typically about 3–5% of the Δ^9-THC content; however, in ditchweed this ratio increases to more than 10%, even though this type of cannabis preparation has the lowest overall mean CBG content. This is because ditchweed has very low Δ^9-THC content (0.4% ± 0.3%). THCV, an important biomarker in cannabis (42,43), is generally present at about 0.5–2.5% of the Δ^9-THC content.

Conclusions

The question over the increase in potency of cannabis is complex and has evoked many opinions. The issue has been clouded somewhat by reports of 10- and 30-fold increases in cannabis potency since the 1970s. It is however clear that cannabis has changed during the past four decades. It is now possible to mass produce plants with potencies inconceivable when concerted monitoring efforts started 40 years ago. The PM program has strived to answer this cannabis potency question, while realizing that the data collected in this and other programs have some scientific and statistical shortcomings. These include randomness of samples, correctly identifying the various cannabis products, sampling, natural degradation of Δ^9-THC over time, and different analytical techniques, making comparing results between countries and over time very difficult. However, analysis of the available data in conjunction with the PM program results makes a strong case that cannabis is not only more potent than in the past but also that this high-potency product's market share is also growing. This is clearly evident in the increase in sinsemilla seizures and in the increase in marijuana and sinsemilla samples with Δ^9-THC >9%. The question now becomes: What are the effects of the availability of high-potency products on cannabis users?

Acknowledgments

The authors appreciate the cooperation of the DEA regional laboratories in submitting all the nondomestic samples, and the states' eradication programs for submitting the domestic samples. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

1. Leggett T. A review of the world cannabis situation. United Nations Office On Drugs and Crime (UNODC). Bull Narc 2006;58:1–155.
2. UNODC World Drug Report 2008, http://www.unodc.org/documents/wdr/WDR_2008/WDR_2008_eng_web.pdf (accessed April 14, 2009).
3. Pacifico D, Miselli F, Micheler M, Carboni A, Ranalli P, Mandolino G. Genetics and marker-assisted selection of the chemotype in *Cannabis sativa* L. Mol Breed 2006;17:257–68.
4. Pacifico D, Miselli F, Carboni A, Moschella A, Mandolino G. Time course of cannabinoid accumulation and chemotype development during the growth of *Cannabis sativa* L. Euphytica 2008;160:231–40.
5. Bócsa I, Máthé P, Hangyel L. Effect of nitrogen on tetrahydrocannabinol (THC) content in hemp (*Cannabis sativa* L.) leaves at different positions. J Int Hemp Assoc 1997;4:78–9.
6. Lewis R, Ward S, Johnson R, Burns DT. Distribution of the principal cannabinoids within bars of compressed cannabis resin. Anal Chim Acta 2005;538:399–405.
7. ElSohly MA, Slade D. Chemical constituents of marijuana: the complex mixture of natural cannabinoids. Life Sci 2005;78:539–48.
8. Appendino G, Giana A, Gibbons S, Maffei M, Gnavi G, Grassi G, et al. A polar cannabinoid from *Cannabis sativa* var. Carma. Nat Prod Commun 2008;3:1977–80.
9. Radwan MM, ElSohly MA, Slade D, Ahmed SA, Wilson L, El-Alfy AT, et al. Non-cannabinoid constituents from a high potency *Cannabis sativa* variety. Phytochemistry 2008;69:2627–33.
10. Ahmed SA, Ross SA, Slade D, Radwan MM, Khan IA, ElSohly MA. Structure determination and absolute configuration of cannabichromanone derivatives from high potency *Cannabis sativa*. Tetrahedron Lett 2008;49:6050–3.
11. Radwan MM, Ross SA, Slade D, Ahmed SA, Zulfiqar F, ElSohly MA. Isolation and characterization of new cannabis constituents from a high potency variety. Planta Med 2008;74:267–72.
12. Ahmed SA, Ross SA, Slade D, Radwan MM, Zulfiqar F, ElSohly MA. Cannabinoid ester constituents from high-potency *Cannabis sativa*. J Nat Prod 2008;71:536–42.
13. Radwan MM, ElSohly MA, Slade D, Ahmed SA, Khan IA, Ross SA. Biologically active cannabinoids from high potency *Cannabis sativa*. J Nat Prod 2009;72:906–11.
14. Mackie K. Understanding cannabinoid psychoactivity with mouse genetic models. PLoS Biol 2007;5:2106–8.
15. Mackie K. From active ingredients to the discovery of the targets: the cannabinoid receptors. Chem Biodiversity 2007;4:1693–706.
16. Galal AM, Slade D, Gul W, El-Alfy AT, Ferreira D, ElSohly MA. Naturally occurring and related synthetic cannabinoids and their potential therapeutic applications. Recent Pat CNS Drug Discovery 2009;4:112–36.
17. Pertwee RG. The diverse CB1 and CB2 receptor pharmacology of three plant cannabinoids: Δ^9-tetrahydrocannabinol, cannabidiol and Δ^9-tetrahydrocannabivarin. Br J Pharmacol 2008;153:199–215.
18. Mechoulam R, Peters M, Murillo-Rodriguez E, Hanus LO. Cannabidiol—recent advances. Chem Biodiversity 2007;4:1678–92.
19. Ramaekers JG, Kauert G, van Ruitenbeek P, Theunissen EL, Schneider E, Moeller MR. High-potency marijuana impairs executive function and inhibitory motor control. Neuropsychopharmacology 2006;31:2296–303.
20. Zuardi AW. History of cannabis as a medicine: a review. Rev Bras Psiquiatr (Braz J Psychiatry) 2006;28:153–7.
21. McLaren J, Swift W, Dillon P, Allsop S. Cannabis potency and contamination: a review of the literature. Addiction 2008;103:1100–9.
22. Smith N. High potency cannabis: the forgotten variable. Addiction 2005;100:1558–60.
23. Carter GT, Mirken B. Medical marijuana: politics trumps science at the FDA. MedGenMed 2006;8:46.
24. Swami M. Cannabis and cancer link. Nat Rev Cancer 2009;9:148.
25. ElSohly MA, Ross SA, Mehmedic Z, Arafat R, Yi B, Banahan BF. Potency trends of Δ^9-THC and other cannabinoids in confiscated marijuana from 1980–1997. J Forensic Sci 2000;45:24–30.
26. ElSohly MA. Marijuana: constituents and potency trends. 219th ACS National Meeting, Book of Abstracts 2000;TOXI:006.
27. ElSohly MA, Holley JH, Turner CE. Constituents of *Cannabis sativa* L. XXVI. The Δ^9-tetrahydrocannabinol content of confiscated marijuana, 1974–1983. In: Harvey DJ, Paton W, Nahas GG, editors. Marijuana '84: proceedings of the Oxford symposium on cannabis. England: IRL Press, 1985;37–42.
28. ElSohly MA, Holley JH, Lewis GS, Russell MH, Turner CE. Constituents of *Cannabis sativa* L. XXIV: the potency of confiscated marijuana, hashish, and hash oil over a ten-year period. J Forensic Sci 1984;29:500–14.
29. Potter DJ, Clark P, Brown MB. Potency of Δ^9-THC and other cannabinoids in cannabis in England in 2005: implications for psychoactivity and pharmacology. J Forensic Sci 2008;53:90–4.
30. Lopes de Oliveira G, Voloch MH, Sztulman GB, Negrini Neto O, Yonamine M. Cannabinoid contents in cannabis products seized in São Paulo, Brazil, 2006–2007. Forensic Toxicol 2008;26:31–5.
31. Pijlman FTA, Rigter SM, Hoek J, Goldschmidt HMJ, Niesink RJM. Strong increase in total Δ^9-THC in cannabis preparations sold in Dutch coffee shops. Addict Biol 2005;10:171–80.
32. King LA, Carpentier C, Griffiths P. An overview of cannabis potency in Europe. Luxembourg, Belgium. European Monitoring Centre for Drugs and Drug Addiction (EMCDDA) 2004. http://www.emcdda.europa.eu/attachements.
cfm/att_33985_EN_Insight6.pdf (accessed April 18, 2009).
33. van Laar M, Cruts G, van Gageldonk A, Croes E, van Ooyen-Houben M, Meijer R, et al. The Netherlands drug situation 2007: report to the EMCDDA by the reitox national focal point. Utrecht, Netherlands. Trimbos Institute, Netherlands Institute of Mental Health and Addiction,

2008, http://english.wodc.nl/images/The%20Netherlands%20Drug%20Situation%202007_tcm45-115793.pdf (accessed April 18, 2009).
34. Niesink RJM, Rigter S, Hoek J, Goldschmidt H. THC concentrations in marijuana, nederwiet and hash in Netherlands coffee shops (2006–2007). Utrecht, The Netherlands: Trimbos Institute, Netherlands Institute of Mental Health and Addiction, 2007.
35. Licata M, Verri P, Beduschi G. Δ^9-THC content in illicit cannabis products over the period 1997–2004 (first four months). Ann Ist Super Sanità 2005;41:483–5.
36. Poulsen HA, Sutherland GJ. The potency of cannabis in New Zealand from 1976 to 1996. Sci Justice 2000;40:171–6.
37. Hall W, Swift W. The THC content of cannabis in Australia: evidence and implications. Aust NZ J Public Health 2000;24:503–8.
38. Holler JM, Bosy TZ, Dunkley CS, Levine B, Past MR, Jacobs A. Δ^9-Tetrahydrocannabinol content of commercially available hemp products. J Anal Toxicol 2008;32:428–32.
39. King LA, Carpentier C, Griffiths P. Cannabis potency in Europe. Addiction 2005;100:884–6.
40. Barnett V, Lewis T. Outliers in statistical data, 3rd edn. New York, NY: John Wiley & Sons Ltd., 1994.
41. Ross SA, ElSohly MA. CBN and Δ^9-THC concentration ratio as an indicator of the age of stored marijuana samples. Bull Narc 1997–1998;49–50:139–47.
42. ElSohly MA, DeWit H, Wachtel SR, Feng S, Murphy TP. Δ^9-Tetrahydrocannabivarin as a marker for the ingestion of marijuana versus Marinol: results of a clinical study. J Anal Toxicol 2001;25:565–71.
43. ElSohly MA, Feng S, Murphy TP, Ross SA, Nimrod A, Mehmedic Z, et al. Δ^9-Tetrahydrocannabivarin (Δ^9-THCV) as a marker for the ingestion of cannabis versus Marinol. J Anal Toxicol 1999;23:222–4.

Additional information and reprint requests:
Mahmoud A. ElSohly, Ph.D.
National Center for Natural Products Research
School of Pharmacy
University of Mississippi
University, MS 38677
E-mail: melsohly@olemiss.edu

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01406.x
Available online at: interscience.wiley.com

PAPER

CRIMINALISTICS

Nicole L. Ledbetter,[1] *M.S.; Barbara L. Walton,*[2] *B.S.; Pedro Davila,*[1] *B.S.; William D. Hoffmann,*[1] *B.S.; Richard N. Ernest,*[3] *B.S.; and Guido F. Verbeck IV,*[1] *Ph.D.*

Nanomanipulation-Coupled Nanospray Mass Spectrometry Applied to the Extraction and Analysis of Trace Analytes Found on Fibers*

ABSTRACT: This article presents the novel instrumentation of nanomanipulation coupled to nanospray ionization-mass spectrometry, which is used to directly probe trace analytes found on individual fibers. The low detection limits and sample volumes associated with nanospray ionization-mass spectrometry make it the ideal instrument to implement for trace analysis. Nanospray ionization-mass spectrometry, coupled with the nanomanipulator, allows for the direct probing of trace particulates on fibers. The technique is demonstrated by dissolving an electrostatic particle of cocaine from a fiber, collecting the analyte solution in a nanospray tip, and transferring the tip directly to the mass spectrometer to complete analysis. The utility of this technique is evident through the minimal sample preparation and short analysis time. The use of nanomanipulation coupled to nanospray ionization-mass spectrometry could improve on current trace particulate analysis by reducing both detection limits and sample size required to complete analysis.

KEYWORDS: forensic science, nanospray, trace analysis, fiber, nanomanipulation, mass spectrometry

Direct probing from a sample surface directly coupled to mass spectrometry (MS) is a useful tool, helping to eliminate sample preparation and analysis time. Currently, there are three techniques at the forefront of direct-coupled surface sampling MS: desorption electrospray ionization (DESI) (1), surface sampling probe electrospray ionization (2), and dielectric barrier discharge ionization source (DBDI) (3). DESI sprays charged solvent droplets onto an ambient surface which ionizes neutral analytes. The analytes are then desorbed from the surface and analyzed using MS (1,4). DESI has been used to detect trace amounts of explosives as well as sampling directly from human skin (5,6). Surface sampling probe electrospray ionization uses a liquid junction between the electrospray source and the surface to dissolve and then ionize the analyte, which is then electrosprayed into the MS (2,7). This method has been used to sample drugs directly from thin tissue slices (8). DBDI uses a dielectric barrier discharge to create a stable plasma flow that desorbs and ionizes the sample off of an ambient surface, then analyzes it using MS (3). All of these techniques have great utility, but need a relatively large area (20–100 μm^2) for analysis.

Micromanipulation is a significant tool in the biological and chemical sciences. It is utilized primarily to manipulate small particles and cellular materials because of its precise movements. It is currently being used in the biological sciences for single cell transfer (9), to isolate specific bacterial cells from a group (10), and it has also been employed for sample preparation for MS analysis (11). Mitochondria have been extracted from cells using micromanipulation and subsequently analyzed using electrophoresis (12). For the purposes of this discussion, nanomanipulation will generally refer to the use of a commercially available instrument from Zyvex (Richardson, TX) that has the capability of manipulating samples as well as extracting target analytes from those samples. As the nanomanipulator was designed to be utilized with electron microscopy, the manipulator endeffectors have better than 5 nm translational resolution, which is beyond the optical limit. This allows for new advances in the biological and chemical sciences to be made through precise movements and minimally invasive sample manipulation.

One of the current methods of probing trace analytes is the swab method, whereby an object's surface is swabbed using a textile sampling swab that is then put into a solvent to extract the analyte of interest (13). This method is not the best way to collect trace analytes because of analyte losses and dilution of analyte concentration. Repetitive handling of the analyte, associated with multistep processes, can lead to sample contamination (14). Additional difficulties arise when attempting to swab a single fiber, as analyte concentrations will be very low, making analysis difficult. Improvement in trace analyte sampling is needed to more accurately solve problems and collect trace evidence.

Mass spectrometry is a useful tool for trace analysis because of its high sensitivity, allowing it to be useful for a wide variety of compounds. Nanospray is an ideal ionization source to couple to nanomanipulation, because it reduces sample preparation time and

[1]Department of Chemistry, University of North Texas, 1155 Union Circle, #305070, Denton, TX 76203.
[2]Department of Chemistry and Biochemistry, Central Connecticut State University, 1615 Stanley St., New Britain, CT 06050.
[3]Department of Biological Science, University of North Texas, 1155 Union Circle, #305220, Denton, TX 76203.

*This research has been generously funded by: NSF-REU CHE-0648843, Zyvex Industrial Grant GN0001371, and UNT Faculty Research Grant G34267.

Received 20 Sept. 2008; and in revised form 15 June 2009; accepted 4 July 2009.

requires a small concentration of analyte (pmol/μL). Nanospray is an ionization technique that, at best, requires 300 attograms (10^{-18} g) of analyte with a minimum volume of 300 nL. Additionally, it is not as affected by salts as electrospray ionization, which further reduces sample preparation (15). Liquid chromatography–electrospray ionization–mass spectrometry (LC–ESI–MS) has been used in the analysis of explosives (16) and other trace analytes because of the ability to deconvolute a large sample matrix. Using nanomanipulation, the nanospray tip is brought to the analyte to discriminate particle selection, which would help to both deconvolute the spectra and elucidate the identity of the analyte. These techniques can be applied to trace analysis, expanding current abilities, so that the trace is now able to be extracted and analyzed.

Materials

The solvents and chemicals utilized were chloroform ($CHCl_3$), glacial acetic acid (HOAc), Optima* LC/MS methanol (MeOH), and caffeine (Thermo Fisher Scientific Inc., Waltham, MA); no further purification was necessary. A sample of freebase cocaine was provided by the University of North Texas Police Department (Denton, TX). Millipore water was obtained using the Milli-Q^{UF} Plus (Millipore, Billerica, MA) with better than 18 MΩ salt content. Glass-bottom dishes were used to hold our samples (Mat Tek Corp., Ashland, MA), and analytes were probed from 100% rayon white bemberg lining. The MS utilized was an LCQ DECA XP Plus (Thermo Finnigan, San Jose, CA) with a nanospray ionization source (Proxeon Biosystems, Odense, Denmark). An L200 nanomanipulator (Zyvex, Richardson, TX), coupled to a TE2000U Microscope (Nikon, Melville, NJ) and a PE2000b four-channel pressure injector (MicroData Instrument Inc., S. Plainfield, NJ) were used to retrieve the analyte from the fiber.

Methods

The L-200 nanomanipulator is mounted to a Nikon TE2000U inverted microscope. The nanomanipulator employs four nanopositioners that can be controlled using a joystick and/or digital input (Fig. 1*a*). The nanopositioners have two modes of action that allow for precise control of their movements. In the coarse mode of action, the nanopositioners have a range of motion of 12 mm in the *X* and *Z* axes and 28 mm in the *Y* axis. The fine mode of action allows for a range of motion of 100 μm in the *X* and *Z* axes and 10 μm in the *Y* axis. The nanopositioners are further distinguished by the type of manipulation tools they use. Two nanopositioners are capable of holding endeffectors (either tungsten probes or microgrippers) that can be utilized in either the coarse or fine mode with 3.4 nm translational resolution. The remaining two nanopositioners, which are run in coarse mode only, hold capillary tips and are capable of 100 nm translational resolution (Fig. 1*b*). The PE2000b pressure injector is used to supply up to 60 psi of injection pressure and 24 inches Hg of fill vacuum to the capillaries, allowing us to retrieve the analyte of interest. The capability of the nanomanipulator to hold up to eight nanopositioners is beneficial because it allows one to conduct multiple probes simultaneously and thus, increases the instrument's capabilities and efficiency.

The Au/Pd-plated nanospray tips were loaded with an appropriate solvent, and then the tip was broken using the nanospray source head. A blank was run to determine any solvent contamination, and a background spectrum of the solvent was taken. The tip was then transferred to the nanomanipulator for trace analyte probing from a rayon fiber that was doped with the analyte of interest and

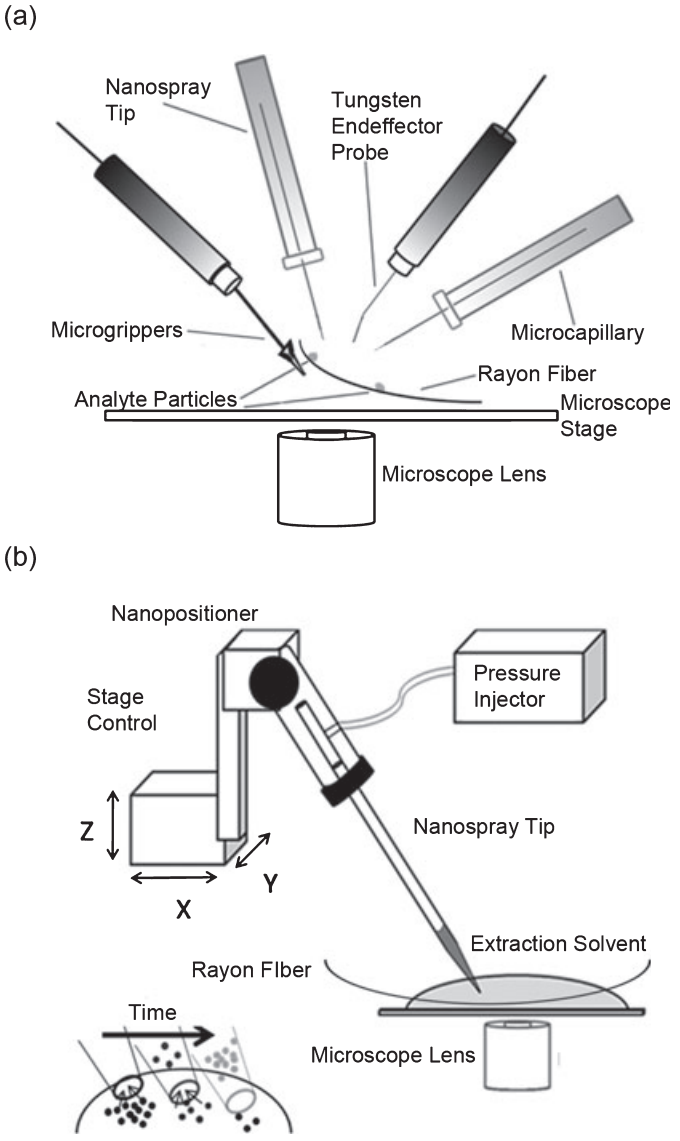


FIG. 1—*(a) Schematic of the nanomanipulator workstation. Two nanopositioners are capable of holding capillary tips used for nanospray ionization. The remaining two positioners utilize endeffectors (either tungsten probes or microgrippers) for sample manipulation. (b) Schematic of a nanopositoner holding a capillary tip for nanospray ionization. Here the extraction solvent is shown on the rayon fiber. A time-resolved schematic shows the retrieval of analyte particles into the tip.*

placed in a glass-bottom dish. The rayon fiber was tacked down to minimize the movement of the fiber and, therefore, the movement of the analyte on the fiber. The particle of interest was found on the fiber, and then the nanospray tip was landed near it, <1 μm away. The nanospray tip then injected the solvent onto the analyte. After the analyte had dissolved, the solvent/analyte solution was retrieved back into the tip. The nanospray tip was then transferred directly to the nanospray ionization source and the sample analyzed. Figure 2*a* shows one of the positioners of the nanomanipulator with a nanospray tip retrieving an analyte. Figure 2*b* shows the tip mounted onto the nanospray ionization source.

Cocaine was used to illustrate this technique. When sampling, 50:50 MeOH: H_2O with 1% HOAc was used as the solvent and 3 μL was loaded into the nanospray tip. The tip was landed next to the analyte as shown in Fig. 3*a*. The nanomanipulator used an injection pressure of 20.8 psi for a duration of 11 ms delivered

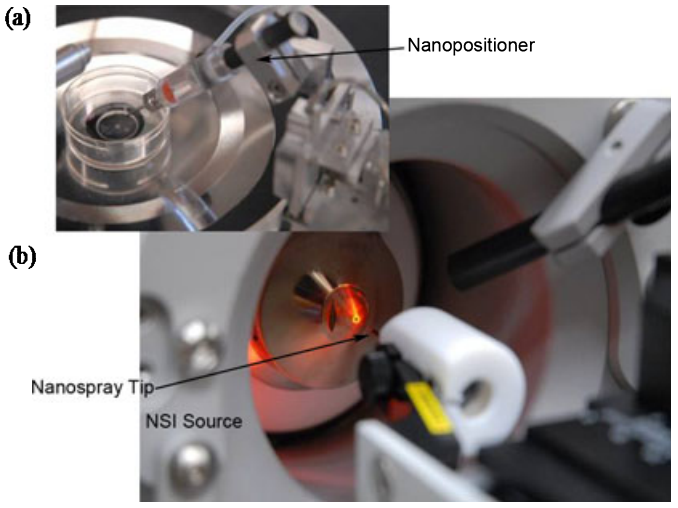


FIG. 2—*(a) The nanomanipulator positioner with the nanospray tip probing an analyte. (b) The nanospray ionization source showing the nanospray tip that was transferred directly from the nanomanipulator.*

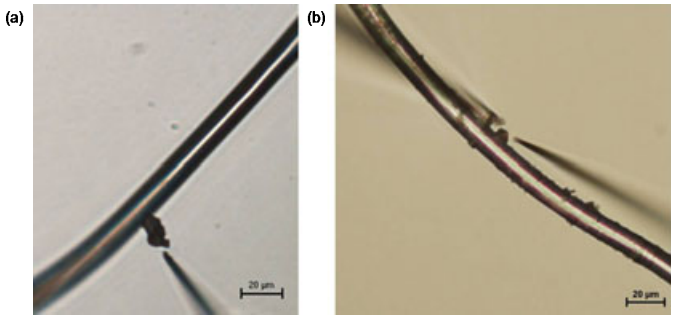


FIG. 3—*The rayon fiber doped with analyte before extraction. (a) The caffeine particle is on the rayon fiber and the nanospray tip is landed in close proximity. (b) Demonstration of the two capillary method with a histidine particle between the capillary tip on the left and the nanospray tip on the right.*

from the pressure injector and a fill pressure of 65.0 psi with a fill time of 50 ms. The sample was then analyzed in the positive ion mode using a 2 kV extraction voltage on the NSI-MS. The mass spectrum of cocaine can be seen in Fig. 4.

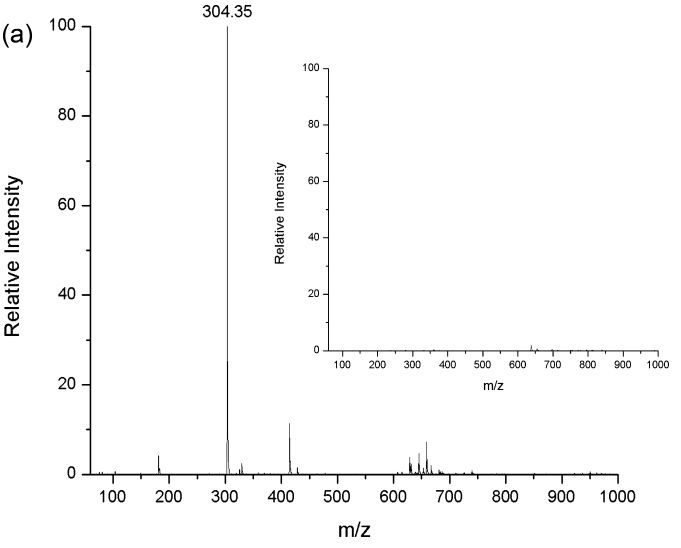


FIG. 4—*(a) Mass spectrum of cocaine (MW: 303.35 g/mol) taken in positive ion mode after extraction from a single rayon fiber. The MH^+ peak appears at $^m/_z$ 304.35. The inset shows the mass spectrum collected from an extraction on the fiber with no analyte present. (b) Blow-up of mass spectrum shown in (a) between $^m/_z$ 250–350.*

Results and Discussion

Cocaine trace particles were sampled directly from a single rayon fiber using the nanomanipulator and then analyzed using NSI-MS. Figure 4*a* shows the mass spectrum of cocaine after directly probing a trace particle from the rayon fiber. As analysis was completed by NSI, the MH^+ peak is most prominent and appears at $^m/_z$ 304.35, as well as the characteristic fragment peak at $^m/_z$ 181.94. The inset of Fig. 4*a* displays the blank when the solvent was allowed to extract on the fiber with no analyte present. As can be seen from the inset, the fiber contributed no appreciable peaks to the mass spectrum of cocaine. Figure 4*b* displays a blow-up of the $^m/_z$ 250–350 range.

Our results clearly show that the nanomanipulator coupled to NSI-MS is an effective instrument to probe trace analytes from fibers. It is an improvement in trace analyte probing from a single fiber, allowing for new experimental procedures to be created and smaller amounts of analyte to be sampled. The nanomanipulator reduces cost of sampling from fibers because of the minimal sample preparation and the reduced analysis time. Computer encoding of the positioners may be implemented to automate the procedure. Being able to recover trace analytes from a single fiber allows for better analysis of crime scenes, and the reduced sample size and volume required for NSI-MS allows for the ability to retrieve a higher sample concentration. Although the research for this article has focused on cocaine, this technique has also been applied to other particulate analyte standards including caffeine and histidine, and the limit of detection by NSI-MS is on the order of 7 pg (10^{-12} g) for histidine, which demonstrates the extreme sensitivity achievable by this technique. This technique has also been applied to analysis of single organelles and protein extraction from silicon beads, currently being developed in our group.

It is important to have a solvent to dissolve the sample as well as provide a steady spray flow. Some nonpolar compounds will not dissolve in any of the solvents appropriate for nanospray, so it is

important to utilize a two capillary system with one capillary containing a solvent to dissolve the analyte, and the other capillary containing the nanospray solvent that retrieves the dissolved analyte. Figure 3*b* illustrates the two capillary system used with some nonpolar analytes. Diffusion will occur, and the small amount of nonpolar solvent with the analyte of interest will mix with the nanospray solvent, and the resulting solvent/analyte mixture can be analyzed using the NSI-MS. The nanomanipulator is also capable of liquid–liquid phase microextractions to sample trace analytes and gain higher sample concentration from a dilute analyte in a liquid sample and then complete analysis.

This research centered around the analysis of particulates on fibers which are particularly well suited for analysis by an inverted microscope but that by no means limits this method to fibers. The technique presented here would work equally well for hair, paper, money, or any other surface where trace particulate analytes are expected. Difficulties arise with some of these specimens because they are not all equally suited to examination by an inverted microscope. In this case, the nanomanipulator stage could be transferred to a noninverted microscope.

Conclusion

Here, we have introduced the novel instrument of nanomanipulation coupled to NSI-MS as an effective tool to analyze trace analytes found on fibers, with the successful analysis of cocaine. We recovered trace particles of cocaine from rayon fibers using the nanomanipulator and subsequently analyzed the trace by directly taking the sample from the nanomanipulator to the NSI-MS. This clearly demonstrated the functionality and utility of the nanomanipulator coupled to NSI-MS to improve upon the current methods of analysis of trace analytes found on fibers.

In the field of trace analysis, preconcentration techniques produce only positive results for metal chelates; therefore, a better option might be a preconcentration technique like evaporation, extraction methodologies, or changing the sampling procedure to obtain a more concentrated sample (17). Advantages of NSI-MS include having the capability to analyze samples with limited volumes (as low as 300 nL) and low limits of detection and would eliminate the need for any such preconcentration technique.

Acknowledgments

We thank LT. West Gilbreath with the UNT police, and Patrick Horn.

References

1. Cooks RG, Ouyang Z, Takats Z, Wiseman JM. Ambient mass spectrometry. Science 2006;311(5767):1566–70.
2. Van Berkel GJ, Sanchez AD, Quirke JME. Thin-layer chromatography and electrospray mass spectrometry coupled using a surface sampling probe. Anal Chem 2002;74(24):6216–23.
3. Na N, Zhao M, Zhang S, Yang C, Zhang X. Development of a dielectric barrier discharge ion source for ambient mass spectrometry. J Am Soc Mass Spectrom 2007;18(10):1859–62.
4. Wiseman JM, Laughlin BC. Desorption Electrospray Ionization (DESI) mass spectrometry: a brief introduction and overview. Curr Sep 2007;22(1):11–4.
5. Takats Z, Cotte-Rodriguez I, Talaty N, Chen H, Cooks RG. Direct, trace level detection of explosives on ambient surfaces by desorption electrospray ionization mass spectrometry. Chem Commun (Camb) 2005;(15): 1950–2.
6. Justes DR, Talaty N, Cotte-Rodriguez I, Cooks RG. Detection of explosives on skin using ambient ionization mass spectrometry. Chem Commun (Camb) 2007;(21):2142–4.
7. Asano KG, Ford MJ, Tomkins BA, Van Berkel GJ. Self-aspirating atmospheric pressure chemical ionization source for direct sampling of analytes on surfaces and in liquid solutions. Rapid Commun Mass Spectrom 2005;19(16):2305–12.
8. Van Berkel GJ, Kertesz V, Koeplinger KA, Vavrek M, Kong A-NT. Liquid microjunction surface sampling probe electrospray mass spectrometry for detection of drugs and metabolites in thin tissue sections. J Mass Spectrom 2008;43(4):500–8.
9. Yamamura S, Kishi H, Tokimitsu Y, Kondo S, Honda R, Rao SR, et al. Single-cell microarray for analyzing cellular response. Anal Chem 2005;77(24):8050–6.
10. Ishøy T, Kvist T, Westermann P, Ahring B. An improved method for single cell isolation of prokaryotes from meso-, thermo- and hyperthermophilic environments using micromanipulation. Appl Microbiol Biotechnol 2006;69(5):510–4.
11. Kajiyama Si, Harada K, Fukusaki E, Kobayashi A. Single cell-based analysis of torenia petal pigments by a combination of ArF excimer laser micro sampling and nano-high performance liquid chromatography (HPLC)-mass spectrometry. J Biosci Bioeng 2006;102(6):575–8.
12. Ahmadzadeh H, Johnson RD, Thompson L, Arriaga EA. Direct sampling from muscle cross sections for electrophoretic analysis of individual mitochondria. Anal Chem 2004;76(2):315–21.
13. Dahl DB, Cayton JC, Lott PF. Gunshot residue analysis: an applicability study. Microchem J 1987;35(3):360–4.
14. Lloyd JBF, King RM. One-pot processing of swabs for organic explosives and firearms residue traces. J Forensic Sci 1990;35(4):956–9.
15. Higbee DJ, Douglas R, Smith J, White TP, Bigwarfe PM, Dolan AR. Miniaturization of electrospray ionization mass spectrometry. Trends Appl Spectros 2002;4:141–54.
16. Vigneau O, Machuron-Mandard X. A LC-MS method allowing the analysis of HMX and RDX present at the picogram level in natural aqueous samples without a concentration step. Talanta 2009;77(5):1609–13.
17. Bernal Morales E, Revilla Vázquez AL. Simultaneous determination of inorganic and organic gunshot residues by capillary electrophoresis. J Chromatogr A 2004;1061(2):225–33.

Additional information and reprint requests:
Guido F. Verbeck IV, Ph.D.
University of North Texas
Department of Chemistry
P.O. Box 305070
Denton, TX 76203
E-mail: gverbeck@unt.edu

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01438.x
Available online at: interscience.wiley.com

PAPER

CRIMINALISTICS

Wei Chu,[1,2] *Ph.D.; John Song,*[1] *M.S.; Theodore Vorburger,*[1] *Ph.D.; and Susan Ballou,*[1] *M.S.*

Striation Density for Predicting the Identifiability of Fired Bullets with Automated Inspection Systems*

ABSTRACT: Automated firearms identification systems will correlate a reference bullet with all evidence bullets without a selection procedure to exclude the bullets having insufficient bullet identifying signature. Correlations that include such bullets increase the workload and may affect the correlation accuracy. In this article, a parameter called *striation density* is proposed for determining and predicting bullet identifiability. After image preprocessing, edge detection and filtering techniques are used to extract the edges of striation marks, the resulting binary image distinctly shows the amount and distribution of striation marks. Then striation density is calculated for determining the quality of images. In the experiment, striation densities for six lands of 48 bullets fired from 12 gun barrels of six manufactures are calculated. Statistical results show strong relation between striation density and identification rate. It can provide firearms identification systems with a quantitative criterion to assess whether there are sufficient striae for reliable bullet identification.

KEYWORDS: edge detection, firearms identification, forensic science, identifiability, morphology, striation density

Traditionally, comparisons of striation marks on fired bullets have been accomplished manually by examiners using comparison microscopes. Since the early 1990s, computer-aided optical systems have been developed to aid firearms examiners by processing large volumes of firearms-related evidence more efficiently. Currently, the Integrated Ballistics Identification System (IBIS^{TM}) developed by Forensic Technology, Inc. (FTI, Cote-Saint-Luc, QC, Canada) (1) is widely used in U.S. forensic laboratories, and is the current standard technology in the National Integrated Ballistic Information Network (NIBIN). Developments of other commercial systems and studies in firearms identification have also been reported (2–7). IBIS and other systems adopt the procedure that ranks the bullets stored in the database in light of a certain similarity metric with a subject bullet. The ranking of the bullets in the database gives the firearms examiner an ordered list of the most promising potential matches for further manual comparison and to decide whether any compared bullets match the subject bullet.

The striation marks on the surface of the fired bullet are formed when the bullet is forced down the firearm barrel. Because the bullet signatures are topographic striations by nature, the quality of the bullet image depends on the quantity of clear striation marks contained in the image. Sometimes, there are only a few, if any, effective striation marks transferred from the barrel to the land impression (also called *land engraved area* or LEA) of the fired bullet. In this case, performing a correlation with this LEA will unnecessarily increase the processing time and may affect the accuracy of a correlation score. This correlation score is often calculated as the sum of correlation scores from each pair of LEAs when the two bullets are at their optimum phase orientation. Although a skillful examiner could judge the image quality based on his/her experience, it would still be difficult to decide whether or not to abandon any further correlation operation because of the lack of an accepted quantitative criterion.

In this article, a new parameter called striation density, d_s, is proposed as an objective criterion for quantifying the suitability of the bullet images for automatic bullet signature correlation. First, we introduce the image preprocessing. Then we describe the edge detection procedure and extraction of effective striation marks from the binary edge image. The evaluation criterion of the bullet image quality is proposed, and then we will discuss this new concept.

Image Preprocessing

The earlier version of IBIS (now known as IBIS Heritage) (8), acquires two-dimensional (2D) image data or photographs of bullet surfaces. The recent version of IBIS uses a three-dimensional (3D) topography imaging system (9) for acquisition of bullets. The National Institute of Standards and Technology (NIST) developed a correlation approach based on 3D topography measurements using commercial confocal microscopy. The approach was originally developed for signature measurements and inspection of the NIST Standard Reference Material (SRM) bullets and cartridge cases (2,10), but it has been adapted for the measurements and correlations of 2D and 3D bullet and cartridge case signatures.

[1]National Institute of Standards and Technology, Gaithersburg, MD 20899.
[2]Harbin Institute of Technology, Harbin, 150001, China.

*The funding for this research is provided by the National Institute of Justice (NIJ) through the Office of Law Enforcement Standards (OLES) at NIST. Certain commercial equipment, instruments, or materials are identified in this paper to specify adequately the experimental procedure. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

Received 27 Mar. 2009; and in revised form 23 July 2009; accepted 4 Aug. 2009.

Theoretically, the extraction of binary edge image and establishment of an identifiability criterion do not rely on the data format acquired by the imaging system, but different image preprocessing may be required for 2D and 3D data. The preliminary image processing at NIST for data acquired by the 3D confocal topography measurement system consists of three steps.

Removal of Unreliable Data Points

The unreliable data points include dropouts and outliers. Some 3D imaging systems provide a value of "level of confidence" associated with each acquired data set. If the level of confidence is too low, the point is considered as a dropout. In contrast with dropouts, "outliers" are data points inaccurately measured by the 3D imaging system, which are not reported to the user as inaccurate by the acquisition hardware (11). In our algorithm, the outliers are detected by estimating the local slope between a point and its neighbors. If the slope is above a threshold of 50 degrees, generally related to the numerical aperture of the objective lens, the slope is regarded to be unphysical and the point is identified as an outlier. After all dropout and outlier points are identified, they are replaced by interpolated data.

Gaussian Filter

After removal of unreliable data points, the topography data can be considered as a grayscale image. The grayscale image contains both low-frequency curvature and high-frequency noise. Both are not considered as *individual characteristics* and must be removed before correlation (11). In this project, the Gaussian filter (12,13) with 0.25-mm long wavelength cutoff λ_c and 0.0078-mm short wavelength cutoff λ_s is used to remove the low-frequency waviness and high-frequency noise. With Gaussian filter processing, topography components with a frequency range out of the specified bandwidth are strongly attenuated.

Top-Hat Transform

Some remnant of the bullet surface curvature may still exist even after Gaussian filtering. In the next step, a top-hat transform is applied to eliminate the remaining curvature. The top-hat transform is one of a large collection of techniques known as mathematical morphology (14), a branch of nonlinear image processing and analysis that concentrates on the geometric structure within an image. Mathematical morphology was initially developed for binary images and later extended to grayscale images. It includes a large number of operators and techniques for binary and/or grayscale image: dilation, erosion, opening, closing, thinning, skeletonization, hit-or-miss transform, and others. The labeling and area open operator will be used in the next section for edge element filtering.

The top-hat transform is a morphological operator to detect subtle structures. Puente León applied this technique in his research on bullet identification (7). Depending on whether bright or dark structures are to be detected, two different transforms, called the open top-hat transform and the close top-hat transform, may be applied. The open top-hat transform, $f \hat{\circ} g$, is given by:

$$f \hat{\circ} g = f - (f \circ g) \tag{1}$$

The close top-hat transform, $f \hat{\bullet} g$, is given by:

$$f \hat{\bullet} g = (f \bullet g) - f \tag{2}$$

where f and g are the image to be transformed and the structuring element, respectively; and the symbols "$\circ$"and "$\bullet$" denote the morphologic opening and closing operators, respectively. If the difference between them is calculated, a signal containing only fine peaks will be obtained, while the coarser structures are suppressed.

Figure 1 shows a 3D bullet topography image acquired by a confocal microscopy (*a*) and the consecutively processed results after removing unreliable data points and interpolation (*b*); Gaussian filter (*c*); and top-hat transform (*d*). During the top-hat transform procedure, a threshold is also set to remove the areas apparently higher or lower than other areas because they are not considered as striation marks.

Edge Detection and Extraction of Striation Information

An edge detection technique is used for extraction of striation information. Edge detection is a process that primarily measures, detects, and localizes changes of intensity (15). It has the desirable property of drastically reducing the amount of information to be processed subsequently while preserving information about the shapes of objects in a scene (16). Edges may or may not correspond to the boundary of an object. For images that do not contain concrete objects, the interpretation of the image depends on its texture properties. The edge detection procedure helps in the extraction of texture properties (17). In a bullet surface image, the textures are mainly comprised of striation marks. The properties of edge detection provide a possibility to quantitatively analyze the striation marks, or bullet signatures, which are of interest to effective bullet identification.

The edge detection is used to localize the striation edges. There are many methods to perform edge detection. We used the Canny edge detector, which is a multi-stage algorithm to detect a wide range of edges in images (18). Figure 2 shows a primary edge detection result using the Canny detector. The pixel points judged to be edges by the detector are evaluated "1" and shown as bright, while other pixel points judged to be background are evaluated "0" and drawn as dark. We define a parameter called edge density, d_e,

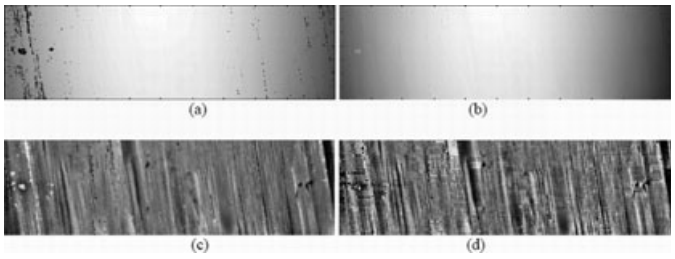


FIG. 1—*Preliminary processing result for a bullet LEA image: (a) raw data; (b) after removal of unreliable data points and interpolation; (c) after Gaussian filter; (d) after top-hat transform. The bullet nose is toward the top of the page.*

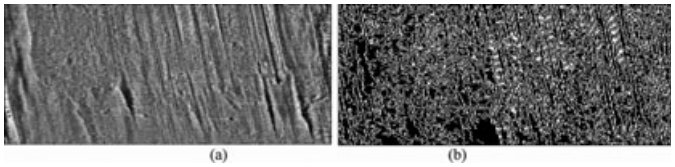


FIG. 2—*Edge detection: (a) grayscale image after preliminary processing; (b) detected result using the Canny detector. The edge density is* d_e = 18%.

to describe the ratio of the number of edge pixel points to the total pixel points of the area:

$$d_e = \text{number of pixels at the edges} / \text{total number of pixels in the area} \quad (3)$$

As indicated in Fig. 2*b*, the edge density is d_e = 18.0%.

The primary edge image depicts the contour of all detected features on the bullet surface, including useful features associated with the striation marks impressed by the barrel, and useless features located within the image areas not engraved by the barrel rifling. In this article, we define a pixel point evaluated "1" by the edge detector as an edge element. An edge element belonging to a striation mark is called a striation element, and the edge elements belonging to all other features are called disturbance elements. In order to extract the striation elements from the disturbance elements, an additional filtering process has been designed.

First, we need to analyze the connection relationship of all the pixel points. Two pixels are said to be 4-neighbors if they are vertically or horizontally adjacent. They are said to be 8-neighbors if they are 4-neighbors or diagonally adjacent. In Fig. 3, all pixels labeled with "+" are 4-neighbors of the center pixel labeled with "•". All 4-neighbors of the center pixel and the remaining pixels labeled with "*" are its 8-neighbors. A collection of edge elements is considered 8-connected if for any two pixels p and q in the collection there exists a sequence of pixels also in the collection such that the first pixel is p, the last is q, and each pixel in the sequence is an 8-neighbor of the next. We define such an 8-connected collection separated from other collections as a connection unit. A morphological labeling operation (14) is then used to identify all connection units and mark them with sequence numbers.

Figure 4 shows three connection units with typical shapes. Fig. 4(*a*) is an edge detection result from a striation mark. It has the property of an approximately straight and continuous long span in the vertical direction. Contrarily, the edge detection results originating from other features irrelevant to striations may have various shapes, straight or tortuous, simple or complicated, fragmentary or extensive, but few of them have a long vertical span, like Fig. 4(*a*) representing a striation mark, or bullet signature. Figure 4(*b,c*) show two examples of edge patterns other than striation marks. In term of this difference, they can be filtered out from the primary edge image.

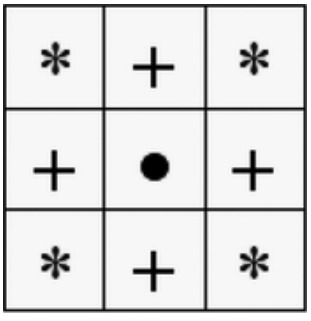


FIG. 3—*Connection relationship.*

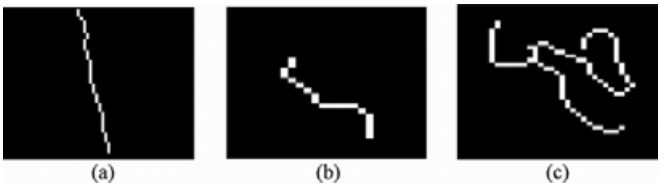


FIG. 4—*Three entity examples existing in the primary edge image. (a) represents a striation mark, or bullet signature; (b) and (c) do not.*

First, a threshold, denoted as l, is used to distinguish them. The vertical span of each connection unit is checked and compared with l. All the units, whose vertical pixel position difference between their top elements and their bottom elements is greater than l, are reserved. Otherwise, they are removed by re-evaluating the pixels as "0". Prior to this processing, a morphological area open operator is used to reduce the processing time. This operator can automatically omit the connection units that have a total number of elements less than l, as Fig. 4(*b*) shows. For more complicated cases, where contours of different features connect together, a tracing process is developed to filter out the disturbance elements connected with the striation edges. At first, a rule for tracing is built and shown as Fig. 5(*a*). Assuming that the dot point is the current point being traced, the rule prescribes that the next possible traced point can only be one of three adjacent points labeled by 1, 2 and 3. The numerical sequence is their priority. The priority sequence of 2 and 3 depends on the direction of twist of the barrel. Figure 5(*b*) shows an example of tracing. The tracing starts from the top pixel point of this connection unit. The tracing path is determined in terms of the established rule until the last pixel point is reached. The pixels along this path will be reserved if the vertical span is more than l. The path in Fig. 5(*b*) is marked in bright, while the edge elements that have not been traced are marked in gray. The same process is repeated until all the edge elements have been considered. Figure 6 shows the filtered edge image from Fig. 2(*b*) in which only the contours of striation marks are highlighted.

Note that the tracing process can complete the filtering alone, but the application of the morphological operator can significantly reduce the processing time.

Evaluation of Bullet Identifiability

By analogy with Eq. 3, we define a parameter called striation density, d_s, to describe the percentage of pixel points on striations within the total area of pixel points:

$$d_s = \text{number of pixels on the striations} / \text{total number of pixels in the area} \quad (4)$$

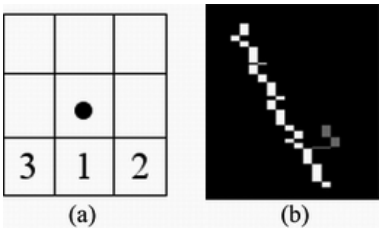


FIG. 5—*Element tracing: (a) rule and priority; (b) a tracing example.*

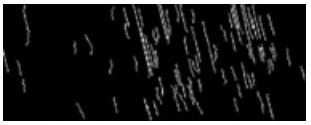


FIG. 6—*Filtered edge image result with a striation density* d_s = 2.90%.

As indicated in Fig. 6, the striation density is $d_s = 2.90\%$. It is obvious that d_s must be less than d_e.

Two more bullet images with edge detection and filtering processing are shown in Fig. 7. Although the edge density d_e for these two sets of images is about the same, 16.9% for (b) and 17.8% for (e), the striation density d_s shows a large difference, 5.70% for (c) and 0.16% for (f). Apparently, the striation density d_s appropriately captures the amount of striations in the bullet image. The more and clearer striation marks a bullet image has, the higher the striation density will be. Striation marks are the foundation for bullet signature correlation, and they determine the reliability of bullet identification. The striation density parameter d_s can provide examiners with a means to estimate the quality of bullet images and predict the potential for identification.

Data from our previous study on correlations among 48 bullets provides us evidence to support the aforementioned point (19). In that study, 48 topographic images (4 bullets × 2 barrels × 6 firearm brands) were selected from a collection of sample bullets fired by a variety of firearms (11,20) and were correlated with each other. An overlap metric P was calculated in order to characterize the quality of the separation (or the degree of overlap) between the matching and nonmatching distributions of correlation values. The parameter P describes the probability that the CCF value of a randomly chosen member from the nonmatching distribution is larger than the CCF value of a randomly chosen member from the matching distribution (20). The parameter P quantifies the potential for making accurate matches. The smaller the value of P the better the separation between matching and nonmatching distributions. A value of zero for the overlap metric P implies ideal identification without error. The data of Figs. 2(*d*), 7(*a*), and 7(*d*) are processed from bullets fired through Taurus, Beretta, and Bryco barrels, respectively. Their image qualities are representative of other bullets fired from same brand barrel as well. Table 1 lists the averaged striation densities of bullets fired from these barrel brands along with the values of P.

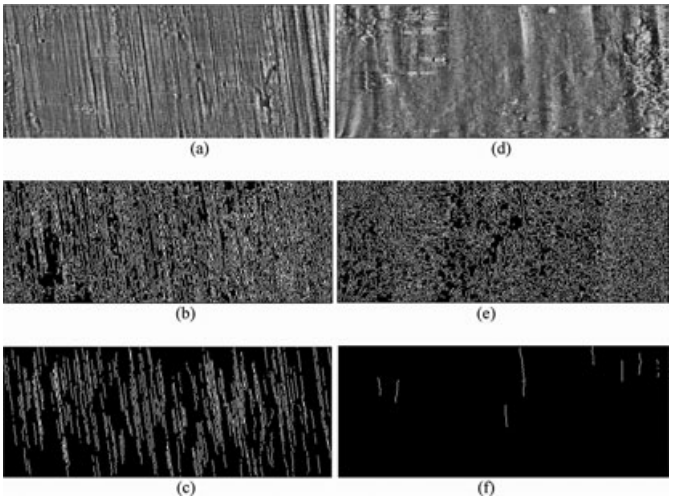


FIG. 7—*Two examples of edge detection: (a) and (d) are two grayscale images after preliminary processing, (b) and (e) are primary edge detection images, and (c) and (f) are filtered edge images. The edge density d_e is 16.9% for (b) and 17.8% for (e), two values which are very close, but the striation density d_s is 5.70% for (c) and 0.16% for (f), indicating a large difference in the image quality for firearms identification.*

TABLE 1—*Averaged striation densities, d_s, of 48 land images of 8 bullets for each of three different barrels.*

No.	Brand	Overlap Metric P	Average Edge Density d_e (%)	Average Striation Density d_s (%)
1	Beretta	0.026	16.5	4.07
2	Taurus	0.277	17.8	1.73
3	Bryco	0.554	16.9	0.84

For bullets fired through the Beretta barrel, the separation is high. Conversely, the overlap metric P of bullets fired through Bryco barrels shows a more random condition. This suggests that the associated striation density at this level is not sufficient for a bullet to be identified. Although a precise relationship between the striation density and the identification rate has not been developed, higher striation densities of compared bullets might lead to higher confidence in correlation results.

Table 1 shows some statistical characteristics of bullets fired from different barrel brands. It does not necessarily mean that bullets fired from certain brand barrels are identifiable or bullets fired from other brand barrels are not identifiable. The result relies on the striation analysis of *individual* bullets and suggests a relationship between the striation density, the image quality and the possible correlation results.

Discussion

In the previous examples, the edge density d_e by itself is not an important factor for bullet identification. Very similar values are calculated from different images. From the numerical experiment, we preliminarily conclude that d_e is not sensitive to image quality when an edge detector is selected. Conversely, the striation density d_s has a strong sensitivity to different image qualities as described in the foregoing sections. More detailed studies for d_e itself and for its relation to d_s are in process to try to reveal more characteristics of bullet signatures.

Sometimes, the striations are not evenly distributed over an entire image. Calculation of the striation density of an entire image might mistakenly cause the exclusion of an LEA image of sufficient quality. For example, for an image with low striation density d_s, a partial area of it may contain strong striation marks and an effective compressed profile can be extracted from it. Manual selection of the area of interest is a straightforward way to solve this problem, but automatic selection by the software would be a better choice for increased objectivity. Figure 8 shows such an example. Figure 8(*a*) is a grayscale image after preliminary image processing. Figure 8(*b*) represents its corresponding striation edge image. The striation density d_s calculated directly from the edge image is 4.08%. If the bottom rectangular area, which does not contain any striation element, is excluded, this value increases slightly to $d_s = 4.18\%$. Further, the useful area can be more strictly selected. Figure 8(*c*) plots the accumulative sum of edge elements of each horizontal section, with the abscissa indicating the accumulative sum of edge elements versus the ordinate indicating the y-direction pixel index. It can be clearly seen that the striation

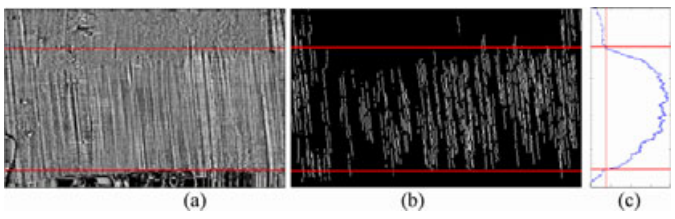


FIG. 8—*Select dense area of striation to recalculate the striation density d_s.*

elements are concentrated at the central part of the edge image. A vertical line further partitions the curve at abscissa values equal to 20% of the maximum sum. Calculating the striation density only from the area lying between the two horizontal lines yields an increased value for $d_s = 5.59\%$.

Distinguishing connection units through a morphological operation makes the analysis of the statistical and individual characteristics of these units possible. All such units after filtering are assumed to be the contours of striation marks in the foregoing section. But we might go beyond this assumption and develop a stricter set of striation selection criteria. Through first-order least-squares fitting, the tilt angle and straightness degree of each striation can be obtained by calculating the slope and the residuals of the fitting line. Thus, we can further decide whether every single striation candidate is reasonable and acceptable by checking whether the tilt angle or straightness degree is within our expectation. However, this is accompanied by an increase in the processing time, and an overly strict selection of striations may also cause a loss of useful information and lead to inaccurate compression of 2D profile data.

Summary and Future Work

Based on edge detection, an objective criterion, known as the striation density d_s, is proposed for determining the bullet identifiability. The relationship between the striation density and the identifiability of fired bullets is verified using identification results obtained with an automated system. However, we believe that the concept can be extended to assess identifiability as well for experts using manual microscope systems. In addition, other derived parameters, such as depth, average length, straightness, or distribution of striations, may also be considered for quantifying the quality of striations.

The concept of striation density is described in this article, but the determination of the threshold for considering striation data to be identifiable will require further research on issues, such as standardization of processing steps and experimental statistical analysis, which in turn will require considerable observation, experimentation and participation by experts.

Acknowledgments

The authors are grateful to B. Bachrach of Intelligent Automation Inc. for confocal image acquisitions of bullets used in this study, to L. Ma of NIST for providing the software code used for dropout processing in the bullet image, and to Bachrach, Ma, and A. Hilton of NIST for helpful discussions.

References

1. Thompson RM. Automated firearms evidence comparison using the Integrated Ballistic Identification System (IBIS). In: Higgins K, editor. Investigation and forensic science technologies. Proceedings of SPIE; 1998 Nov 03;Boston, MA: SPIE, 1999;94–103.
2. Song J, Vorburger TV. Proposed bullet signature comparisons using autocorrelation functions. Proceedings of NCSL; 2000 Jul 16–20; Toronto, Canada. Washington, DC: National Conference of State Legislatures, 2000.
3. Bachrach B. Development of a 3D-based automated firearms evidence comparison system. J Forensic Sci 2002;47(6):1253–64.
4. De Kinder J, Bonifanti M. Automated comparison of bullet striations based on 3D topography. Forensic Sci Int 1999;101(2):85–93.
5. Banno A, Masuda T, Ikeuchi K. Three-dimensional visualization and comparison of impressions on fired bullets. Forensic Sci Int 2004; 140(2–3):233–40.
6. Heizmann M. Automated comparison of striation marks with the system GE/2. In: Geradts ZJ, Rudin LI, editors. Investigative image processing II. Proceedings of SPIE; 2002 Apr 4; Orlando, FL: SPIE, 2002;80–91.
7. Puente León F, Beyerer J. Automatic comparison of striation information on firearm bullets. In: Casasent DP, editor. Intelligent robots and computer vision XVIII: algorithms, techniques, and active vision. Proceedings of SPIE; 1999 Sep 20–21; Boston, MA: SPIE, 1999;266–77.
8. http://www.forensictechnology.com/p7.html (accessed November 25, 2008).
9. Dillon JH. BulletTRAX™-3D, matchpoint Plus™ and the firearms examiner. Evidence Technology Magazine 2005 July-August.
10. Ma L, Song J, Whitenton E, Zheng A, Vorburger TV. NIST bullet signature measurement system for SRM (standard reference material) 2460 standard bullets. J Forensic Sci 2004;49(4):649–59.
11. Bachrach B. A statistical validation of the individuality of guns using 3D images of bullets. Washington, DC: Department of Justice, 2006 Mar; Report No.: 213674.
12. American Society of Mechanical Engineers. ASME B46.1-2002, Surface texture, surface roughness, waviness and lay. New York, NY: American Society of Mechanical Engineers, 2003.
13. Song J, Rubert P, Zheng A, Vorburger TV. Topography measurements for determining the decay factors in surface replication. Meas Sci Technol 2008;19(8):084005.
14. Dougherty ER, Lotufo RA. Morphological processing of gray-scale images. In: Weeks AR Jr, editor. Hands-on morphological image processing. Bellingham, WA: SPIE Press, 2003; 129–62.
15. Poggio T, Voorhees H, Yuille A. A regularized solution to edge detection. J Complex 1988;4(2):106–23.
16. Sarkar S, Boyer KL. On optimal infinite impulse response edge detection filters. IEEE Trans Pattern Anal Mach Intell 1991;13(11):1154–71.
17. Huang JS, Tseng DH. Statistical theory of edge detection. Comput Vis Graph Image Process 1988;43(3):337–46.
18. Canny J. A computational approach to edge detection. IEEE Trans Pattern Anal Mach Intell 1986;8(6):679–98.
19. Chu W, Song J, Vorburger T, Yen J, Baulou S, Bachrach B. Pilot study of automated bullet signature identification based on topography measurements and correlations. J Forensic Sci 2010;55(2):341–7.
20. Vorburger T, Yen J, Bachrach B, Renegar TB, Filliben JJ, Ma L. Surface topography analysis for a feasibility assessment of a national ballistics imaging database. Gaithersburg, MD: National Institute of Standard and Technology, 2007 May; Report No.: NISTIR 7362.

Additional information and reprint requests:
Wei Chu, Ph.D.
Guest Researcher
National Institute of Standards and Technology
100 Bureau Drive, Stop 8212
Gaithersburg, MD 20899-8212
E-mail: wei.chu@nist.gov

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01453.x
Available online at: interscience.wiley.com

PAPER

DIGITAL & MULTIMEDIA SCIENCES

Minhua Ma,[1] *Ph.D.; Huiru Zheng,*[2] *Ph.D.; and Harjinder Lallie,*[1] *M.Sc.*

Virtual Reality and 3D Animation in Forensic Visualization*

ABSTRACT: Computer-generated three-dimensional (3D) animation is an ideal media to accurately visualize crime or accident scenes to the viewers and in the courtrooms. Based upon factual data, forensic animations can reproduce the scene and demonstrate the activity at various points in time. The use of computer animation techniques to reconstruct crime scenes is beginning to replace the traditional illustrations, photographs, and verbal descriptions, and is becoming popular in today's forensics. This article integrates work in the areas of 3D graphics, computer vision, motion tracking, natural language processing, and forensic computing, to investigate the state-of-the-art in forensic visualization. It identifies and reviews areas where new applications of 3D digital technologies and artificial intelligence could be used to enhance particular phases of forensic visualization to create 3D models and animations automatically and quickly. Having discussed the relationships between major crime types and level-of-detail in corresponding forensic animations, we recognized that high level-of-detail animation involving human characters, which is appropriate for many major crime types but has had limited use in courtrooms, could be useful for crime investigation.

KEYWORDS: forensic science, virtual reality, augmented reality, 3D animation, forensic computing, forensic visualization, scene reconstruction, natural language processing, computer vision, motion tracking

Computer-generated animation is an ideal media to accurately visualize crime or accident scenes to the viewer to help understand the situation and retain complex spatial information. Computer-related crime does not involve physical motion in the same way as for instance a homicide; it would, therefore, serve little purpose for such a crime to be reconstructed visually. Computer-generated animation has a limited role to play in the reconstruction of computer-related crime and computer forensics; at most it may be restricted to the demonstration of technical data for the education of juries. For instance, the operation of hard disks and the manner in which they store data can be visualized for judge and jury rather than presented them with a technical description. The problem with such an approach, however, is that the computer-based animations must be acceptable in a courtroom situation.

Based upon factual data, forensic animations can reproduce the scene and demonstrate the activity and location of vehicles, objects, and involved persons at various points in time. Once the animation has been produced, it is easy and cost-effective to observe the scene from various viewpoints such as a driver's view, a victim's view, and a witness' view. Using computer animation techniques to reconstruct crime scenes is replacing the traditional illustrations, photographs, and verbal descriptions, and it is becoming popular in today's forensics.

Computer graphics technology has been successfully applied in forensic visualization, ranging from traffic accident reconstruction to major crime scenes (1–5). Computer-generated (CG) animation has been used to investigate crimes and shown to the jury in the courtroom for evidence presentation (1).

This article integrates 3D graphics, computer vision, motion tracking, natural language visualization, and forensic computing, to investigate the state-of-the-art in forensic 3D animation. The rest of article is organized as the follows: First, section CG Animation for Forensic Visualization discusses approaches of applying CG technologies to forensic reconstruction in mixed reality and virtual reality; Second, Technologies for Forensic Visualization, followed by a description of Crime Types and Level-of-Detail (LOD), and finally the article is concluded by Summary and Discussion.

CG Animation for Forensic Visualization

There are four approaches of adopting computer-generated animation in forensic reconstruction: virtual reality (VR), augmented reality (AR), CG 3D animation, and combining real and synthetic imagery.

Virtual reality simulation in forensic process starts from modeling 3D virtual objects and humans based on measurements and photos and animates the models to recreate the crime scene or incidents concerned. Applications of VR in forensic animation include pathological visualization, murder reconstruction, and shooting case briefing tool.

Augmented reality is the combination of VR and real-world content where CG virtual objects or humans are superimposed over real objects or into video footage in real time. An AR user may wear translucent goggles, through which he could see the real world as well as CG images projected on top of that world. Applications of AR in forensics include simulating road traffic accidents where virtual vehicles are blended into the real scene or

[1]School of Computing, University of Derby, Derby, U.K.
[2]School of Computing & Mathematics, University of Ulster, Co. Antrim, U.K.

*Work supported by the Research Inspired Curriculum Fund 2007/8 from University of Derby, U.K.

Received 14 April 2009; and in revised form 14 July 2009; accepted 23 July 2009.

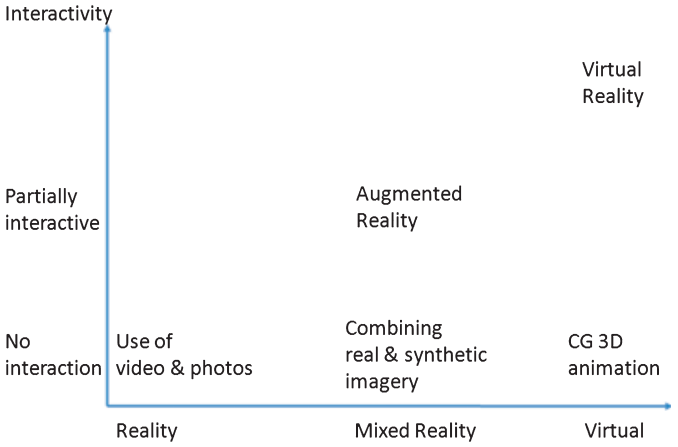


FIG. 1—*Forensic visualization techniques.*

footage and augmented crime scenes where 3D models of victim, suspect, or missing weapon are blended into the real crime scene for evidence presentation. The advantage of using AR in forensic visualization is saving time and costs by reducing the amount of 3D modeling required and higher immersion because of use of real-world elements.

Figure 1 shows the techniques used for forensic visualization. The horizontal axis indicates the proportion of virtual content, and the vertical axis expresses the degree of interactivity. *Use of video, photos, and illustrations* provides no interaction and all the contents are real; therefore, it locates at the bottom left. *CG 3D animation* contains all virtual content and has no interactivity involved when watching the animation, hence it is at the bottom right. *Combining real and synthetic imagery* is somewhere in the middle, and depends on how much synthetic content is used. It has become a popular special effects technique in the movie industry.

The difference between VR/AR and the other three techniques in Fig. 1 is interactivity and real time. VR and AR provide an interactive real-time 3D graphical environment that responds to user actions such as moving around the virtual world or maneuvering virtual objects. VR/AR can put the user in the driver's seat for accident reconstructions and allow the user to observe a crime scene from a desired vantage point, which is impossible to film or observe. A VR/AR user can also, for example, play the various roles of victim, perpetrator, or witness to experience the reconstruction of the incident.

In 2003, Linden Lab released Second Life, a virtual world in which users can create and animate their own objects and characters. Object creation is facilitated by the availability of basic prims (i.e., primitives), which can be modified, textured, and scripted. The behaviors of objects and nonplayer characters (bots) are programmed using the Linden scripting language. Second Life has been used for crime scene reconstruction and role-playing for crime investigation, mainly for educational purposes. There are places built in Second Life for role-play in forensic science teaching to provide live forensic science experience, with roles of victim, perpetrator, witness, crime scene investigator, police, prosecutor, and defender (e.g., *Midian City - A Dark RP Community*) and forensic pathology laboratories (e.g., *Medical Examiner's office*). In a project reported in (6), a group of forensic students have constructed a virtual crime scene complete with a fully furnished house, shoreline, and backyard deck with barbeque, which is a part of the story in Second Life. They then planted clues around the crime scene of a stolen sword. Another group of students examines the crime scene, collects clues, runs forensic tests using functioning equipment, and tries to solve the crime.

Though forensic visualization and digital entertainment industry often use the same sets of software and hardware for creating animation or interactive systems, it is important to differentiate between forensic reconstruction and entertainment. The major differences between them are similar to the differences between animation and simulation: accuracy and purpose. What is aesthetically and perceptively real is unnecessarily physically accurate. Take character motion as an example: professional animators usually tweak the timing, joint ranges, and motion shape of characters for aesthetic edits in digital games and 3D movies industry; whereas forensic animators aim to generate kinematic and dynamic motion as close as possible to the reality.

Technologies for Forensic Visualization

Recent developments in computer technology, especially graphics, have created a climate where novel computing forensic applications have emerged. In this section, we identify the areas where new applications of 3D digital technologies and artificial intelligence (AI) could be used to enhance particular phases of forensic visualization to create 3D models and animations automatically and quickly or to provide real-time interactivity of the reconstructed scenes.

Firstly, we identified three types of data involved in the forensic visualization processes: original, conclusive, and user data. Original data obtained through observation and measurement of the scene; conclusive data is calculation results based on analysis of the original data; and user data is solely for evidence presentation purposes.

Motion Tracking

VR and AR use motion tracking technologies to capture users' movement to provide interaction in real time. Zhou and Hu (7) provided a survey of motion tracking technologies, which consistently update spatiotemporal information with regard to human movement. Figure 2 shows the main categories of motion tracking techniques, including sensor-based, camera-based, glove-based tracking, and haptic interface. A VR/AR user usually wears either sensors or markers on his limbs/joints/head to enable the system to capture his movement in the virtual environment.

Sensor-based systems use electromagnet, electromechanical, inertial, or ultrasound technologies to collect precise data of three-dimensional location and orientation by transmitting and sensing signals from the sensors. Camera-based tracking uses video cameras to detect active (LEDs) or passive (coated with a retroreflective material) markers attached at specific body locations to tracking the movements. Some of the camera-based systems even do not need to place markers on user. Data glove systems usually use tilt and flex sensors to capture the bending of user's fingers. Haptic interface is a special motion tracking device that interacts with a user via touch.

Computer Vision Technologies

Besides being a part of motion tracking technologies in camera-based systems, computer vision technologies have been applied for automatic generation of photo-realistic 3D calibrated models from a sequence of images, e.g., instant Scene Modeler (8), an automatic 3D modeling system that creates calibrated photo-realistic 3D models by using a handheld stereo camera for recording images and a laptop for acquisition and processing.

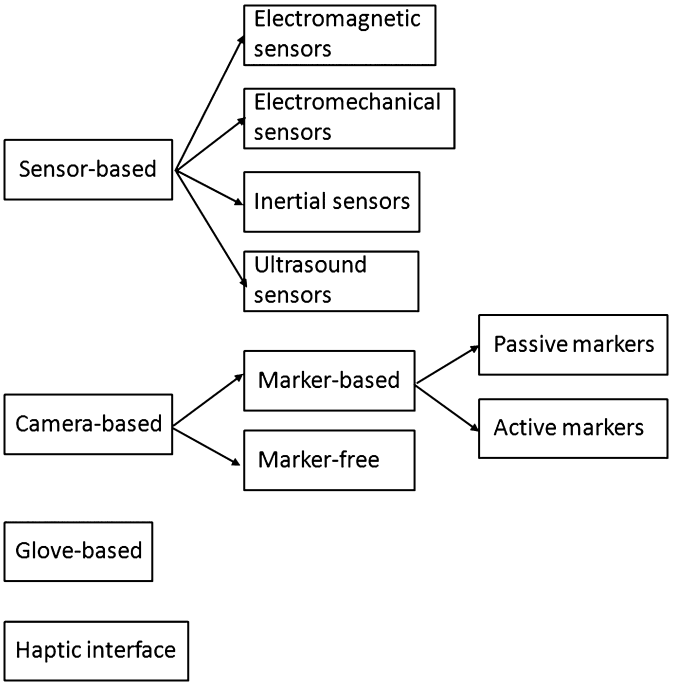


FIG. 2—*Motion tracking techniques.*

Dynamic Simulation

Use of CG visualization in forensic evidence presentation has become admissible in the United Kingdom and globally. In the United Kingdom, computer records are acceptable as evidence if they can be proved to be correct and accurate (1). This leads to the significance of introducing dynamic simulation (a.k.a. physically based simulation) to forensic animation, because dynamic simulation is capable of accurate modeling of objects and human motions based on Newton's laws of motion.

All virtual objects, human kinematics, and motion must obey Newton's laws of motion in a precise simulation. For example, in the case of traffic accident reconstruction, high physical fidelity of the creation of car and victim models is crucial to answer questions about how the accident happened, the sequence of events, and how a specific vehicle moved in the situation. This is the reason that most forensic reconstructionists recognized by courts have been mechanical engineers, and the accepted laws of motion and mechanics form the basis of this expertize.

Currently, a forensic reconstructionist calculates the placement of vehicles before, during, and after an accident based on the data collected from the accident scene and comes up with data representing the movement of all the participants in the accident. An animator then creates 3D models of virtual objects and humans and constructs the scene according to the measurement of the scene, photos, or scale drawings, and data calculated by the forensic reconstructionist.

Dynamic simulation provides realistic motion of virtual objects by modeling the behavior of virtual objects and their responses to external force (e.g., gravity) and torque in a physically realistic manner. Dynamic simulation models objects with their physical properties such as mass, inertia, barycentre, joint limitations, restitution, and surface friction—it can make objects in virtual worlds not only look real but act real. Typical dynamic simulation includes collision detection and the simulation of gravity, friction force, torque, and kinematics in motor actions. Physics can be applied to rigid bodies or deformable bodies such as human tissues.

Real-time dynamic simulation has been used quite heavily in digital games, mechanical simulations, medical visualization and training, and engineering (3). The advantage of applying physically based simulation techniques in forensic visualization is that by virtue of its replayability, adjustability, and reliableness. Physically accurate forensic animation may shed light on investigating what exactly happened at a specific crime scene, causes and effects that embrace the issue of who is at fault or guilty in the case of traffic accidents. In addition, dynamic simulation can be utilized to perform experiments that are impossible or expensive in the real world.

The quality of physics required in a simulation is application specific. In the cognitive and affective domains of learning where the focus of training is more on attitudes, high physical fidelity is not always necessary. However, in medical simulation such as surgery planning and training, high quality physical fidelity is so important that without it the skills acquired in the virtual world may not be transferred to the real one. In forensic animation, the reconstruction of vehicle accidents requires rigid body dynamics and the bombing reconstruction involves breakable rigid body dynamics. The simulation of the biomechanical movement of the human body is an important part of visualization for many major crime types, and some animated pathology sequences may require dynamic simulation of deformable bodies such as muscles. We will discuss this further in Crime Types and Level-of-Detail.

Natural Language Visualization

Natural language visualization is a novel research area that integrates natural language processing (NLP) and 3D animation generation to automate the processes of generating human animation from natural language input (9). Forensic visualization is already benefitting from the outcomes of research to facilitate the reconstruction processes, which is traditionally conducted by police investigators, forensic reconstructionists, and professional animators to create a virtual simulation that fits all the existing facts.

CarSim

CarSim (2) is an automatic text-to-scene conversion system that visualizes car accidents from written reports of motor vehicle accidents. It understands the accident conditions by automatically extracting pieces of information from texts and presenting visually the settings and the movements of the vehicles in 3D scenes. CarSim has been applied to a corpus of French and Swedish texts for which it can currently synthesize visually 35% of the texts. CarSim is also being ported to English.

CarSim represents accidents by applying information extraction techniques to input texts, which reduce the text content to formalized templates that contain road names, road configuration, number of vehicles, and sequence of movements of the vehicles involved. The visualizer reproduces approximately 60% of manually created templates. CarSim's NLP module combines regular expression matching with dependency parsing to carry out the linguistic analysis of the texts. A regular expression grammar is used to identify proper nouns. CarSim focuses on collision verbs, which are vital in the domain of motor vehicle accidents. The visualization module recreates the 3D scene and animates the vehicles. It represents both the entities and the motions symbolically, without taking into account physics laws in the real world.

Human Animation from Natural Language

CONFUCIUS (9) is a framework of language visualization using 3D animation techniques with NLP to achieve high-level animation generation. It is able to visualize single sentences, e.g., *John handed a gun to Nancy*, into 3D animation, speech, and sound effects. The system uses lexical visual semantic representation to connect linguistic semantics to visual semantics of action verbs and to represent verb meanings for action execution (animation). CONFUCIUS gives promising results on word sense disambiguation (70% accuracy) with regard to the data set on which it was tested.

Although originally targeted to application in storytelling, the system has a potential of reproducing crime scenarios based on text descriptions.

Forensic Visualization Processes and Contributions of Computer Technologies

Computer vision technologies automate the processes of obtaining original data and generating 3D models from it. Dynamic simulation automates the process of generating conclusive data from original data. Natural language visualization animates 3D models based on original and conclusive data. Integration of the bespoken technologies provides a potential for automating the processes of reconstructing crime scenarios from evidence that requires minimal intervention from human experts such as crime scene investigators, forensic reconstructionists, and 3D artists.

Figure 3 illustrates the main processes of forensic visualization and where the above-mentioned technologies may contribute to the processes. The double line arrows indicate technology contributions, and the single line arrows indicate information flow between processes.

The data flow from "Data collection" to "Animation" provides physical features of objects and humans that may affect animation, for example, weight and resilience of object, locations after the incident, and ground friction. It also includes the crime types that require different level-of-detail (LOD) in the animation.

The flow from "Data analysis" to "Animation" provides a number of scenarios to be created based on hypothesis. The information flow from "Data collection" to "3D modeling" provides measurement of physical properties of objects and humans that affect creating and scaling of their 3D models, such as size, proportional information, color, texture, and other visual properties.

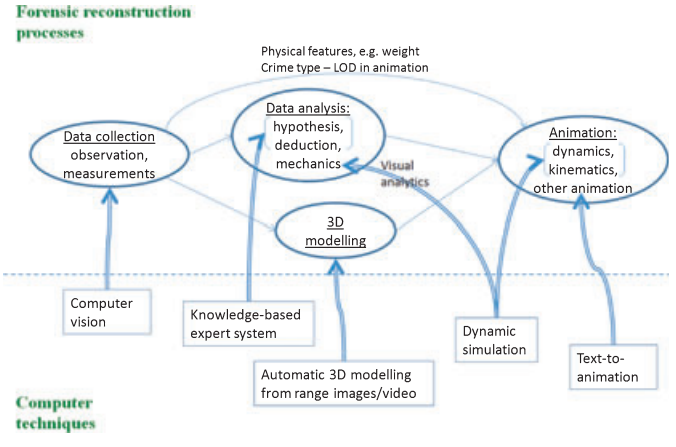


FIG. 3—*Computer technologies contribute to the forensic visualization processes.*

The double line arrows indicate contributions of computer technologies to the forensic visualisation processes. The double line arrow from "Dynamic simulation" to "Data analysis" means performing experiments in dynamic simulation to formulate or rule out a hypothesis and assist deduction, or even discover unexpected situation. As the research only focuses on reconstruction of forensic scenes, knowledge-based expert systems used for deductive reasoning are not discussed in this article.

Crime Types and Level-of-Detail

Level-of-detail (LOD) is a useful concept for managing graphic complexity across many scales. In many VR systems, a virtual object often has multi-resolution representations of polygonal mesh, either refining or simplifying according to certain criteria such as the distance of the object to the camera, so that when it is close to the camera higher LOD mesh is presented, and when it is far away from the camera a low LOD model is used. This has proved an effective optimization strategy in computer graphics. LOD could be supported by having multiple articulations, a.k.a. Levels of Articulation (LOA), when apply to jointed systems like virtual humans. For instance, a low LOA virtual human may be based on a 6-joint skeleton, and a higher LOA virtual human can have 18 or even 71 joints (e.g., H-Anim standard LOA1 and LOA2 as shown in Fig. 4).

H-Anim standard provides four LOAs for applications which require different levels of detail. Some applications such as medical simulation and design evaluation require high fidelity to anthropogeometry and human capabilities, whereas games, training, and visualized living communities are more concerned with real-time performance. Most majority of forensic visualisation contains human figures and is concerned with accurate simulation of

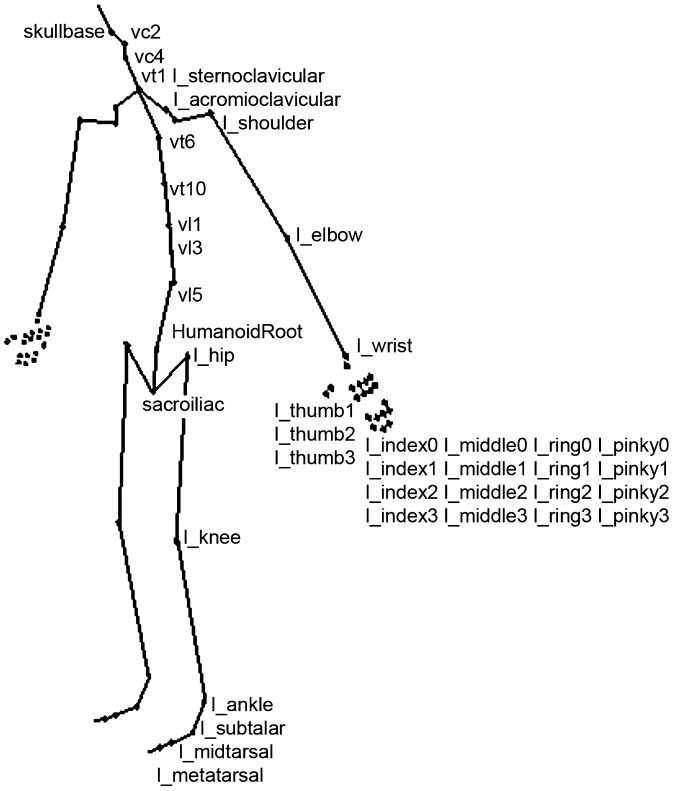


FIG. 4—*Joints and segments of H-Anim LOA2.*

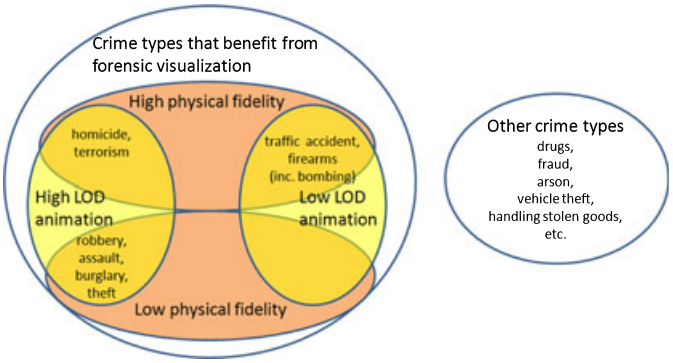


FIG. 5—*LOD, physical fidelity, and crime types.*

humans. Level 2 of Articulation (LOA2) of H-Anim, which has 71 joints, is considered sufficient for human modeling in forensic reconstruction (9). This level ensures enough joints for human movements in forensic visualisation, e.g., it includes enough hand joints for grasp postures. Figure 4 illustrates the joints of LOA2. The dots denote joints and lines segments.

Specific incident reconstruction deals with traffic accidents, bombings, homicides, and accidents of any severity. The requirement of LOD is closely related to the type of crime involved. Figure 5 shows the relationships between main crime types and LOD and physical fidelity needed in the animation.

High physical fidelity and low LOD animation (the overlapped area of the two oval shapes), e.g., traffic accident reconstruction, has been admitted and presented in courtrooms already in the United Kingdom and U.S.A. (10,11). However, high LOD animation involving human figures is rarely admitted in courtrooms, except in the case of pedestrian collision, because of the risk of its overpersuasion power.

Risk of Faulty Interpretation

The issue of creating bias by only watching one scenario of visualization brings the problem of admissibility of displaying CG animation in courtroom. The judge and jury may not be aware of the error or uncertainty involved in measuring and reconstructing the scene, and hence they may be subconsciously biased toward a belief in the presented digital display. According to Lederer and Solomon (12), people are five times as likely to remember something they see and hear rather than something they hear alone; and they are twice as likely to be persuaded if the arguments are buttressed with visual aids.

Despite the many benefits of CG visualization that we have mentioned earlier, it is absolutely essential to verify authenticity, fairness, and relevance before the visualization is used as evidence and presented in court. The original data collected and used in each process (the circles in Fig. 3), the accuracy of the methods (the squares in Fig. 3), and the final visualization must be proved. Another solution to reduce levels of prejudice is visualizing and showing different scenarios of both sides for the defense and for the prosecution in court. How this can be done in practice is out of the scope of this article. It also could be used during the investigation process to analyze and eliminate hypothesis.

Discussion

In summary, CG animation in VR/AR has the ability to communicate highly complex, technical spatial, and temporal evidential information of incident scenes and bring about increased accuracy or speed of the forensic process and can hence reduce the costs involved. The research investigates the state-of-the-art in forensic 3D animation and identifies various technologies that are beginning to be used for or have a potential to be applied in forensic visualization, such as motion tracking, computer vision, dynamic simulation, natural language visualization, and AI. We also analyze the forensic visualization processes and discussed the areas to which these computer technologies may contribute. Having looked at the relationships between major crime types and LODs, it is recognized that high LOD animation involving human characters, which is appropriate for many major crime types, had only a limited use in courtrooms because of admissibility (1), but it could be helpful for crime investigation and informal briefs.

References

1. Burton AM, Schofield D, Goodwin LM. Gate of global perception: forensic graphics for evidence presentation. Proceedings of the 13th Annual ACM International Conference on Multimedia; 2005 Nov 6–11; Singapore. New York, NY: Association for Computing Machinery, 2005;103–11.
2. Egges A, Nijholt A, Nugues P. Generating a 3D simulation of a car accident from a formal description: the CarSim System. In: Giagourta V, Strintzis MG, editors. Proceedings of the International Conference on Augmented, Virtual Environments and Three-Dimensional Imaging; 2001 May 30–June 1; Mykonos, Greece. Thessaloniki, Greece: Informatics and Telematics Institute - CERTH, 2001;220–3.
3. Ma M, McNeill M, McDonough S, Crosbie J, Oliver L. Physics fidelity of virtual reality in motor rehabilitation. In: Richir S, Klinger E, editors. Virtual: a real success. Proceedings of the 8th International Conference on Virtual Reality (VRIC—Laval Virtual 2006); 2006 April 26–30; Laval, France. Laval Cedex, France: Laval Virtual, 2006;35–41.
4. Sauter PM. Introduction to crime scene reconstruction using real-time interactive 3D Technology, http://www.pmsmicro.com/forensicsciences-chapter_4d.pdf (accessed June 11, 2009).
5. Chisum WJ, Turvey BE. Crime reconstruction. London/Burlington/San Diego: Elsevier Academic Press, 2006.
6. Forensic Student Meeting in-world, http://pacificrimx.wordpress.com/2008/06/14/forensic-student-meeting-in-world/ (accessed June 12, 2009).
7. Zhou H, Hu H. Human motion tracking for rehabilitation—a survey. Biomed Signal Process Control 2008;3(1):1–18.
8. Se S, Jasiobedzki P. Instant scene modeler for crime scene reconstruction. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition; 2005 June 20–26; San Diego. Washington, DC: IEEE Computer Society, 2005;123.
9. Ma M. Automatic conversion of natural language to 3D animation [Ph.D. thesis]. Derry, UK: School of Computing & Intelligent Systems, University of Ulster, 2006.
10. Schofield D. Animating and interacting with graphical evidence: bringing courtrooms to life with virtual reconstructions. Proceedings of International Conference on Computer Graphics, Imaging and Visualisation; 2007 Aug 14–17; Bangkok, Thailand. Los Alamitos, CA: IEEE Computer Society, 2007;321–8.
11. Noond J, Schofield D, March J, Evison M. Visualising the scene: computer graphics and evidence presentation. Sci Justice 2002;42(2):89–95.
12. Lederer FI, Solomon SH. Courtroom technology—an introduction to the onrushing future. Fifth National Court Technology Conference (CTC5), National Centre for State Courts, 1997, http://www.ncsconline.org/D_Tech/ctc/showarticle.asp?id=80 (accessed June 11, 2009).

Additional information and reprint requests:
Minhua Ma, Ph.D.
School of Computing
University of Derby
Derby
U.K.
Email: m.ma@derby.ac.uk

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01451.x
Available online at: interscience.wiley.com

PAPER

GENERAL

Robert C. Chisnall,[1,†] *M.Ed.*

Knot-Tying Habits, Tier Handedness, and Experience*

ABSTRACT: Previous research concerning tier handedness, experience and the configuration of simple, habitual knots has been scant and conflicting. Survey data were collected from 21 disparate groups comprising 562 respondents in total. Regardless of experience, respondents tied both Granny and Reef Knots. Dextral tying was dominated by S knots. Left-handers tied Z knots more frequently than right-handers. However, the frequency of S and Z knots relative to tier hand dominance, which is not binary, occurred on a continuum. Averaging all survey tasks, more than 70% of knots tied by dextrals were S, whereas only 56% of those tied by sinistrals were S. These percentages varied somewhat according to specific tying tasks and the number of working ends. Furthermore, tiers' shoelace and parcel knots were not always identical, and a ranked pattern in parcel and shoelace knots was revealed. The examination of habitual knots could benefit criminal investigations.

KEYWORDS: forensic science, behavioral sciences, mirror-image knots, tying habits, handedness, experience, learning

The forensic analysis of knots can be applied to tied ligatures recovered from crime and death scenes, as well as knotted materials involved in occupational and recreational accidents. Regardless of this potential range of useful evidence, knot analysis is a relatively underdeveloped field with few associations to mainstream research and forensic training (1,2). Nevertheless, it can offer a number of corroborating details and useful leads in investigations.

A standard analysis can determine knot identities, functions, locations, tensions, inversions, tying methods, tying order, and likely tier characteristics (1,3). The latter may include habitual and learned tying behavior, the number of tiers involved, and possible tier occupations and hobbies (3,5). At a subtler level, a detailed analysis could distinguish different scenarios—for example, homicide, suicide, and autoerotic fatalities (5,6). Of course any investigation may also rely on the conventional identification of tying materials, trace evidence analysis, and the examination of cordage fibers and cut ends (1,3). Considerable time and effort are required to properly analyze knots, and often the conclusions are equivocal or even useless (1–3).

This is attributable to several key issues. Even though it takes years to become adept at tying, recognizing, and analyzing knots through personal interest and study, there are no widely recognized formal credentials for knot investigators (1). Next, the popular and scientific knotting literature is characterized by conflicting esoterica, such as nomenclature and classification systems. Finally, the practice of forensic knot analysis is founded on knotting lore and tradition. Although forensic knot analysis draws on years of case work and general observation, relevant scientific studies are limited. New cases present unanswered questions, any one of which could occupy years of behavioral research.

Although forensic knot investigation is unlikely ever to be a principal technique, all research undertaken in forensic knot analysis adds to the gradually growing body of knowledge, which could assist in criminal and civil investigations (3). Knot tier handedness relative to habitual knot-tying behavior is just one facet that has received attention over the past three decades. (Whatever influence intrinsic and extrinsic factors may have on this relationship remains undetermined.) This research can assist in understanding the strengths and limitations of interpreting habitual knot tying behavior. It is the subject of this paper.

Fundamental Nomenclature and Structures

Most forensic cases present rudimentary Overhand Knots (Fig. 1), Overhand Loops (Fig. 2), Slip Loops (Fig. 3), Half Knots (Fig. 4), and Half Hitches (Fig. 5), which make up Granny Knots (Figs. 6 and 7) and Reef Knots (Figs. 8 and 9). (The latter are also known as Hercules Knots, or Square Knots in North America.) Overhands and Half Hitches are everyday knots anyone can tie in a variety of combinations. Insofar as this study was concerned, Overhand Knots, Half Hitches, and Half Knots were the main focus. More complex knots that point to training and experience—like the Bowline, Sheet Bend, and the Figure Eight (Figs. 10, 11, and 12, respectively)—are observed less frequently in case work.

In knotter's parlance, the end of the rope or cord is referred to as the working end, or "wend." That portion of the tying material employed to create the knot proper is referred to as the "bight," and the unknotted portion of the cord or rope is called the standing part or "stand" (3,7).

[1]Department of Outdoor and Experiential Education, Faculty of Education, Queen's University, Kingston, ON, Canada.

*Brief references were made to survey data during presentations at Forensic 310, Trends in Forensic Science—The Forensic Analysis of Knots and Ligatures, Trent University, ON, Canada, March 2006, January 2007, 2008, and 2009 and the MAFS Fall Meeting—The Forensic Analysis of Knots and Ligatures, Midwestern Association of Forensic Scientists, Inc., October 12, 2006.

[†]Present Address: 60 Dauphin Avenue, Kingston, ON K7K 6B1, Canada.

Received 22 Feb. 2009; and in revised form 1 Aug. 2009; accepted 9 Aug. 2009.

FIG. 1—*Overhand knots, S (left) and Z (right).*

FIG. 3—*Overhand slip loops, S (left) and Z (right).*

FIG. 2—*Overhand loops, S (left) and Z (right).*

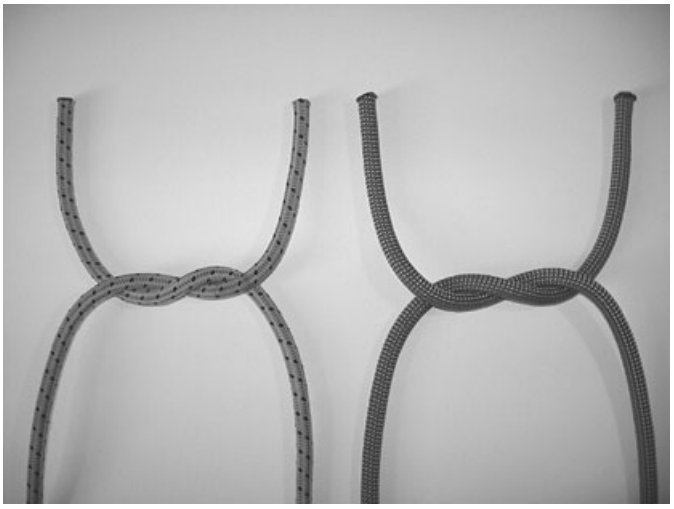


FIG. 4—*Half knots, S (left) and Z (right).*

Although knot references differ on certain aspects of naming and classifying knots, there are some universally accepted categories. A knot or ''stopper knot'' is a compact tangled mass located anywhere in a rope, like an Overhand Knot, Figure Eight, or Stevedore's Knot (Fig. 13). It contains no loops and it is not connected to or tied around any object. A ''bend'' is the joining of two wends, and examples include the Double Fisherman (Fig. 14), which is a proper bend, and the Reef or Square Knot (Figs. 8 and 9), which is a ''packaging knot'' (3). ''Loops'' or loop knots contain one or more fixed loops or movable slip loops. Examples include the Bowline (Fig. 10), which is a fixed loop, and the Overhand Slip Loop (Fig. 3), which is adjustable. A knot tied around a rigid object is referred to as a ''hitch.'' Two common examples are the Girth and Clove Hitches (Figs. 15 and 16, respectively). If the object to which a hitch is tied is removed, the hitch will fall apart or collapse into a stopper knot. This basic classification system does not include splices, braids, and decorative knot work (7).

Some knots, like Overhand Knots and Bowlines, have mirror images and are said to have ''chirality,'' a term chemists use when discussing mirror-image molecules (3,8). (Knot aficionados have liberally borrowed applicable terms from organic chemistry and topology.) Other knots are ''achiral'' or ''amphichiral,'' like the Figure Eight (Figs. 12, 13), because they have no mirror images (1,2). The Figure Eight can distort or ''capsize'' into the Pretzel Knot (Fig. 12), which is also amphichiral. These knots are ''isomorphic'' or topologically equivalent in three-dimensional (3-D) space (8,9). (The term ''homeomorphic'' refers to higher-dimension relationships—as with a 4-D Trefoil and Torus, for example—which are not relevant to practical 3-D knots.)

The two versions of Overhand Knots and Half Hitches are mirror images. One version twists or wraps to the left, the other to the right (Fig. 1). This information is relevant to the study described herein because Overhands and Half Hitches are the basic building blocks of other knots, like Reef Knots and Grannies (Figs. 6 to 9), which also have mirror images. As described by Budworth (4) and Nute (2), the norm has been to call Overhand Knots that twist to the right, right-handed Overhand Knots, and those that twist to the left, left-handed Overhand Knots. This convention matches the way in which hawser-laid ropes are classified: left-hand lay and right-hand lay. However, these labels do not correspond to tier handedness.

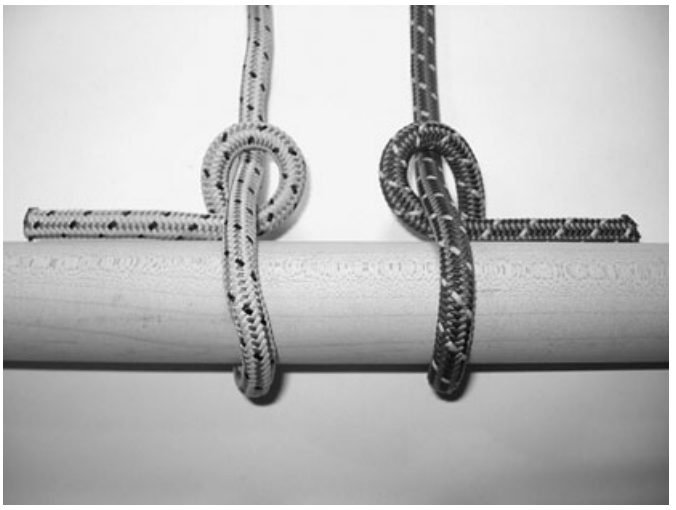


FIG. 5—*Half hitches, S (left) and Z (right).*

FIG. 7—*Z/Z Granny knot.*

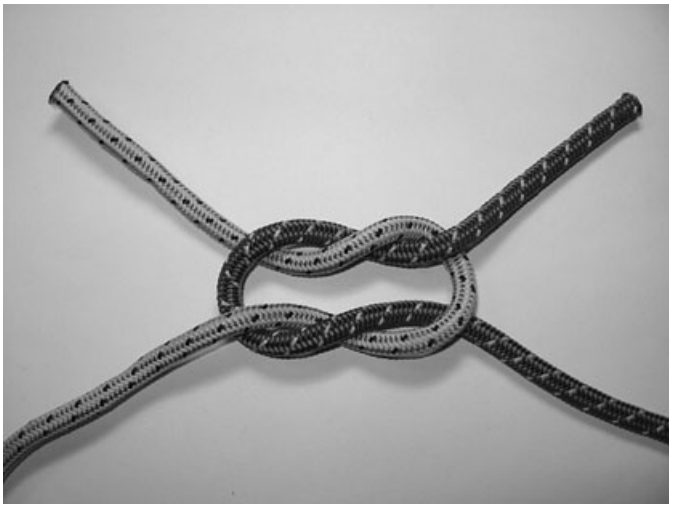


FIG. 6—*S/S Granny knot.*

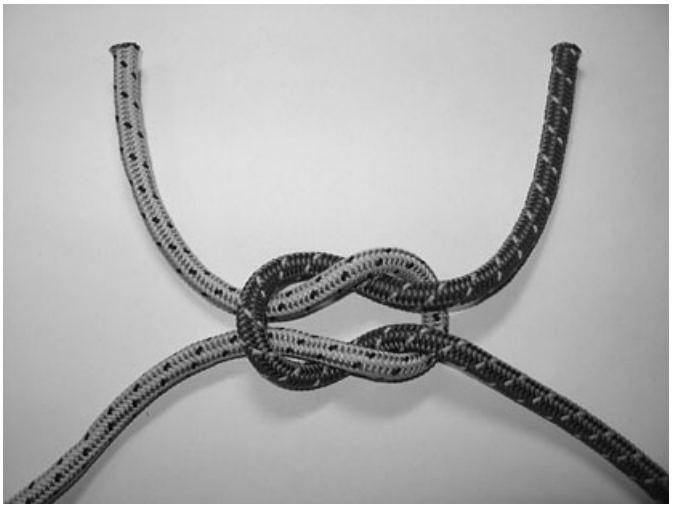


FIG. 8—*Z/S Reef knot (a.k.a. Square or Hercules knot).*

Herein, Overhand Knot "enantiomorphs" or "enantiomers" are designated S or Z (Fig. 1), following government and academic standards for classifying the lay of textile yarns and strands as well as knots (10–12). Several other terms and symbols have been used to denote the direction of twist: RH and LH; R and L; Dextral and Sinistral; Dextra and Sinistra; D and S; D or d (for Dextrorotatory or Dexiotrop) and L or l (for Levorotatory or Laeotrop); and N and Z (3,4,9,12). Bowlines and Sheet Bends, which have similar structures, are assigned the labels S or b and Z or d (Figs. 10, 11).

Granny Knots will be referred to as S/S (Fig. 6) or Z/Z (Fig. 7) configurations, and Reef Knots will be designated as either Z/S (Fig. 8) or S/Z (Fig. 9) configurations. Granny Knots contain two identical Half Hitches, but Reefs require two opposite hitches. The letter in front of the slash—whether S or Z—refers to the first hitch tied. The letter that appears after the slash denotes the Half Hitch closest to the wends, the one tied second. Similarly, Granny bow knots will be labeled S/S (Fig. 17) or Z/Z (Fig. 18), and Reef bow knots will be identified as Z/S (Fig. 19) or S/Z (Fig. 20). The Reef can "capsize" or distort into two opposite Half Hitches forming a Cow or Girth Hitch (Fig. 15), and the Granny can transform into two identical Half Hitches forming a Clove Hitch (Fig. 16).

Why is chirality so important? Note how many distinct combinations are possible with just two simple Half Knots or Half Hitches (Figs. 6–9, 15, 16, 17–20). This does not include the various Grass Knots and Wire Bends that are also formed by two Half Knots (not illustrated). Mirror-image knots should be distinguished when case knots are identified and when habitual knot tying behavior is investigated.

Implications for this Study

It is rare to find people using esoteric or sophisticated knotting techniques. Even those who have specialized training or who are involved in a hobby or occupation that requires unusual knots may not exhibit specialized tying skills. Probably less than 5% of the general population has this kind of training and experience, depending on their geographical location and the prevalent activities (3). The majority of people have very simple, unsophisticated tying habits. They produce simple Overhand Knots and Half Hitches.

The most basic, common knots have mirror images, so it is only natural to ask the following question: Is there any relationship between a tier's handedness and his or her tying habits? Fundamentally, the survey described herein was a quantitative investigation

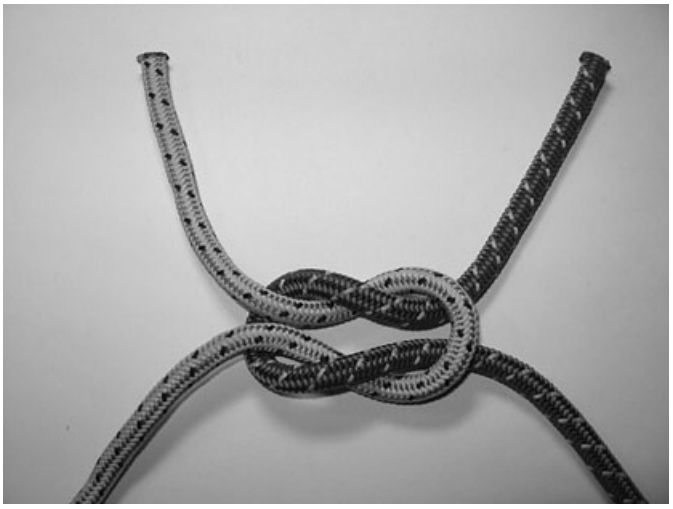


FIG. 9—*S/Z Reef knot.*

FIG. 11—*Sheet bends, S or b (left) and Z or d (right).*

FIG. 10—*Bowlines, S or b (left) and Z or d (right).*

FIG. 12—*Figure eight knot (left) and Pretzel knot (right).*

of this relationship. The four main areas of interest were knot-tying habits, mirror-image knots, tier handedness, and prior knotting experience or learning. How the first three might be related to brain laterality or hemispheric dominance was beyond the realm of this study.

Herein, knot tiers will be designated as sinistrals (left-handers, LH) and dextrals (right-handers, RH). The significance of tier handedness relative to tying habits will become clear as the data are presented.

Methods

The primary aim of this study was to discover if there is any correlation between tier handedness and the chirality of knots tied. If there is an association, what is it and does that finding support or refute previous research? A secondary aim was to acquire some indication of how "literate" the general population is with regard to practical knotcraft, and how acquired knotting skill affects tying habits. Specifically, do people who claim to know the Reef Knot use it when tying shoelaces or parcels? A comparison of shoelaces to parcel knots was the final focus and an attempt to answer the following question: to what extent do tiers produce similar knots when performing these tasks?

Data were collected from 21 different survey groups ranging in size from seven to 139 individuals. Most of the respondents were Canadian citizens residing in southern Ontario at the time, some were U.S. citizens, although the exact number is not known, and 22 volunteers resided in England. Surveys were administered in classrooms, lecture halls, and various meeting venues.

The initial pilot study surveyed 113 subjects utilizing a long questionnaire, which asked respondents to provide the following information: age and gender; knowledge of the Reef or Square Knot, and any previous knot-tying training and experience; writing hand, throwing hand, and the hand used to cut with scissors, brush teeth, strike a match, and input a phone number; which hand performs various tasks during sports and work activities; and kicking foot, dominant eye, and dominant ear. Respondents were then asked to perform the following tasks: tie three Overhand Knots in each of three separate pieces of cord, tie a shoelace knot, and tie a cord around an object. That initial survey was too long and cumbersome, and more tying tasks were needed.

FIG. 13—*Overhand (left), figure eight (middle), and Stevedore's knots (right).*

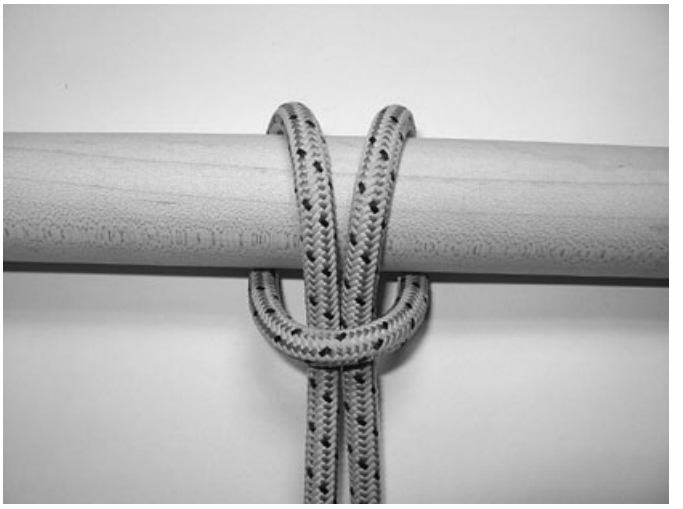


FIG. 15—*Girth hitch (a.k.a. Cow Hitch or Lark's Head).*

FIG. 14—*Double fisherman (example of a bend).*

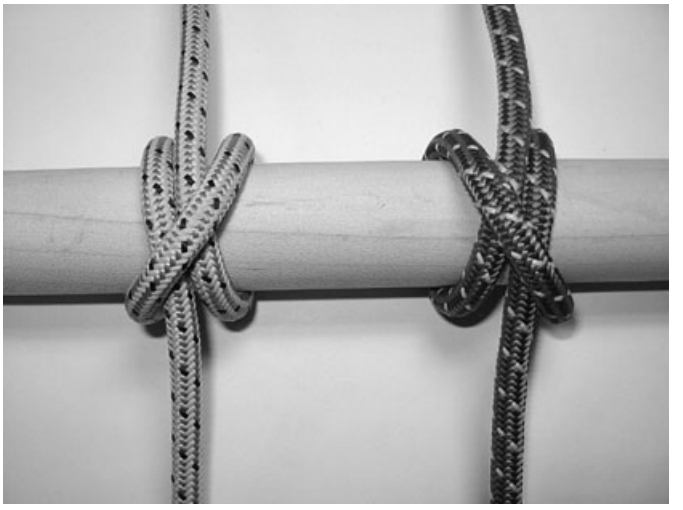


FIG. 16—*Clove hitches, S/S (left) and Z/Z (right).*

The remaining 449 respondents completed a shorter version of the survey instrument. Therein, volunteers were asked to provide the following information: age and gender, writing hand, dominant hand, knowledge of the Reef or Square Knot, and other knotting knowledge. These respondents were asked to perform five knot-tying tasks: tie three Overhand Knots in one cord; tie one Overhand Knot in each of three separate cords; secure a cord around an object as if tying a parcel or package; tie a shoelace knot; and tie four Overhand Knots in a length of cord containing a central pre-tied loop, two knots in each wend. The purpose of that loop was to force tiers to use just one wend when tying the four knots. Respectively, these tasks will be referred to as follows: (i) the one-cord, three-knot task; (ii) the three-cord, three-knot task; (iii) the parceling knot; (iv) the shoelace knot; and (v) the looped-cord, four-knot task.

The provided nylon cords averaged 38 cm in length. They were 2, 3 or 4 mm in diameter, and the wends were cut cleanly and melted. Cord construction was plaited, braided, or double braided. Cabled, twisted, or hawser-laid cord was not employed.

Where indicated, three basic statistical tests were employed to analyze the survey data. The chi-squared test for homogeneity demonstrated that the data could not have been evenly distributed where p-values were less than 0.05. For data rankings, the Spearman Rank Correlation Coefficient and the Wilcox Signed Rank Test were utilized.

Results

A total of 562 volunteer respondents were surveyed between 1988 and 2006 (Table 1). The sample population consisted of mainly volunteer high school students, university undergraduate and graduate students, forensic scientists, and police officers. However, approximately 40 unaffiliated volunteers also participated. There were 339 male and 211 female participants, with 12 volunteers not revealing their gender. No medical limitations were indicated.

The age of the respondents ranged from 5 years to 77 years, with 88% of respondents falling between 10 and 50 years of age. Forty-eight respondents did not reveal their age. Of those reporting their handedness preferences, 477 were dextral writers and 78 were sinistral. Four respondents indicated they were ambidextral or had shared handedness tendencies, and three respondents did not

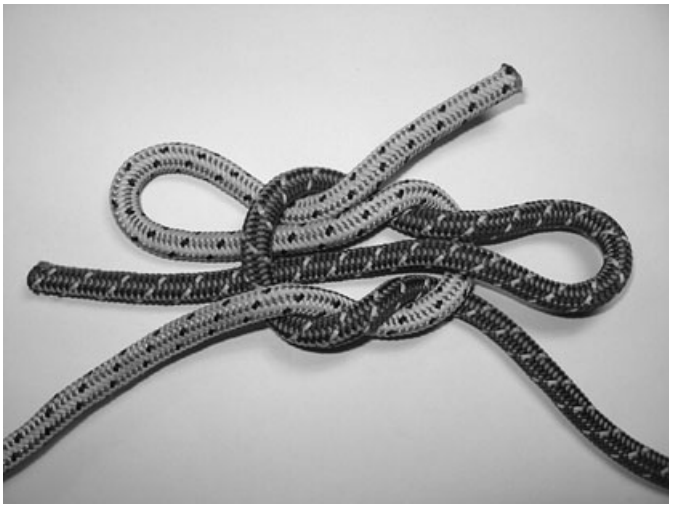


FIG. 17—*S/S granny bow (shoelaces).*

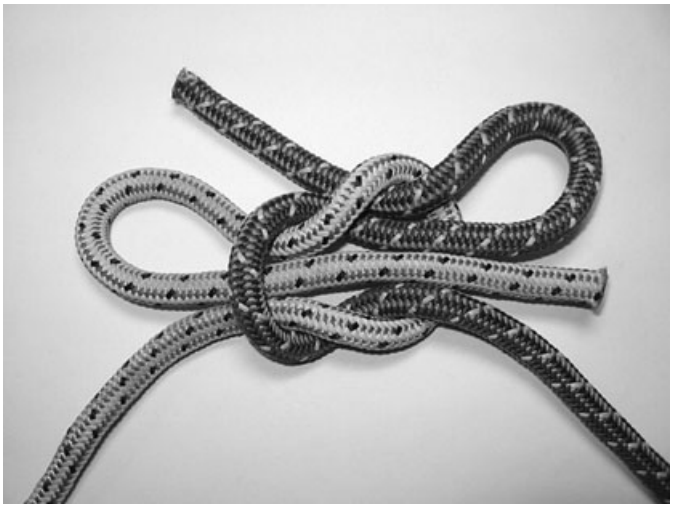


FIG. 19—*Z/S reef bow (shoelaces).*

FIG. 18—*Z/Z granny bow (shoelaces).*

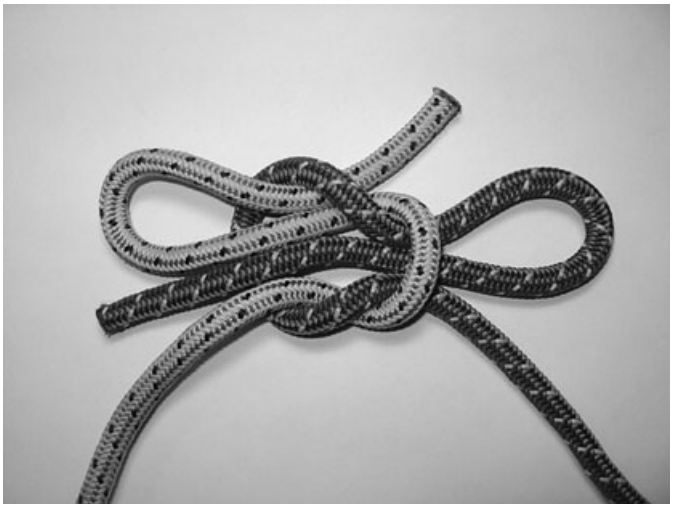


FIG. 20—*S/Z reef bow (shoelaces).*

indicate their handedness. These individuals are not included in any table tallies.

The graduate students and high school students surveyed appeared to have the least interest and background knowledge in knots and related activities. A large number of police officers surveyed claimed to have military, camping, scouting, or sailing experience, all of which involve knots. Many Bachelor of Education students reported having particular interests in outdoor activities. Some of the undergraduate student volunteers were rock climbers and their knotting knowledge was the most extensive of any group surveyed.

Each participant produced, on average, a dozen knots. The minimum number of knots tied by a single subject was zero and the maximum was 15. Approximately 4000 individual knots or knot groups were analyzed. Some participants did not complete each survey task, and some respondents did not answer all survey questions. Several individuals, for certain tasks, exhibited unique tying habits that could not be categorized by general quantitative analyses. Additionally, the short survey did not contain exactly the same questions and tasks as the original long survey. Hence, there are variations in the tabulations presented.

Where indicated, the tables herein summarize the totals of either individual knots, single Half Knots or Half Hitches, knot groups based on task, or the number of survey participants.

Tier Handedness and Knot Chirality

Table 2 summarizes respondent handedness and how frequently respondents habitually tied the Z or S Half Knot or Overhand Knot in all tying tasks. Overall, dextrals tied S Half Knots far more frequently than Z Half Knots. In comparison, sinistral respondents tied S slightly more frequently than Z during all five survey tasks combined. There are several ways in which this data can be analyzed in greater detail.

Utilizing the Half Knot totals for parcel and shoelace tasks alone, as shown in Table 3, it can be seen that 66% of the knots produced by dextrals were S-twisting, which is a trend consistent with Table 2 data, although the S knot is less common among dextrals. Conversely, sinistrals produced the Z enantiomer in 56% of their Half Knots, a trend opposite to that shown in Table 2. For both tables, the chi-squared test for homogeneity indicated the probability that S and Z counts should be 50% is less than 0.001.

TABLE 1—*Survey demographics: gender, handedness, and age.*

	Male (339)			Female (211)			Unknown (12)			
Age	LH	RH		LH	RH		LH	RH		Totals
1–10	1									1
11–20	23	129		10	84			3		249
21–30	9	67	1U	11	75	1A		2		166
31–40	6	26	2A	2	8					44
41–50	4	23	1A	2	3					33
51–60	1	11			4					16
61–70		2			2					4
71–80		1								1
Unknown	6	26		3	6			5	2U	48
Totals	50	285	4	28	182	1		10	2	562

LH, Sinistral; RH, Dextral; A, Ambidextral; U, Unknown.

TABLE 2—*Tier handedness and knot chirality: tabulation of individual Overhand Knots, Half Hitches, and Half Knots from all five tasks.*

Chirality	Knots Tied by Sinistral Tiers	Knots Tied by Dextral Tiers
S	331	2779
	(56% of hitches tied by LH)	(71.5% of hitches tied by RH)
	(10.6% of all S hitches tied)	(89.4% of all S hitches tied)
Z	262	1105
	(44% of hitches tied by LH)	(28.5% of hitches tied by RH)
	(19% of all Z hitches tied)	(81% of all Z hitches tied)

Does the order of tying affect knot chirality? To answer that question, Table 4 shows separate totals for all first and second Half Knots tied during the parcel and shoelace tasks. In the case of dextrals, the S Half Knot predominated during the parcel and shoelace tasks for both the first and second Half Knots. Overall, sinistrals tied Z more frequently regardless of tying order and tying task.

Note the differences in Half Knot frequencies between the two tasks. Regarding dextrals, the more common S enantiomer occurred with a higher frequency in the first Half Knot tied for both tasks, signifying fairly consistent tying habits among those respondents. The tying of the bows or slip loops in the shoelace task did not substantially alter the frequencies. In the case of sinistrals, however, the frequency of the more common Z enantiomer was higher for the first Half Knot in the parceling task and higher for the second Half Knot in the shoelace task, rendering an even split in the totals. This suggests that the tying of slip loops or bows—the second Half Knot in the shoelace task—may cause sinistrals to reverse their habitual enantiomer more often than dextrals.

Do the tallies for the other three survey tasks show a similar trend? Across all three additional tasks, more tiers were disposed to tie the S hitch or knot (Table 5). Again, that tendency was stronger in dextrals, and the chi-squared test for homogeneity indicated the probability is less than 0.001 that S and Z counts in Tables 4 and 5 should be 50%.

What, then, is the difference between these and the parcel and shoelace tasks? The latter tasks utilize two working ends, whereas the other three tasks usually entail the manipulation of one wend only. The use of one wend appears to have driven the average S knot frequency up to 75% for dextrals and 63% for sinistrals.

Instead of counting individual knots, it is valuable to look at a range of tying habits according to tier numbers. The data summarized in Table 6 indicate, first of all, that most respondents did not tie exclusively S or Z knots. They also show that a greater number of dextrals demonstrated tying habits characterized by mostly S rather than mostly Z knots. On the other hand, the number of sinistrals who tied mostly S knots was only slightly higher than those who tie mostly Z knots. The chi-squared test indicated that the probability of an even data distribution is less that 0.001.

The range of knot-tying behavior can be compared to a handedness range. The data summarized in Table 7 are taken from the original long survey, wherein the questions were adequate to determine the degree of dominance or lateralization in each of the 105 respondents completing the entire survey. There were insufficient data to make any conclusions about ambidextrals, mainly right-handed and left-handed respondents. Those who were right-handed for all surveyed tasks produced mostly S knots and hitches. Those who were mainly right-handed, participants who indicated they

TABLE 3—*Half Knot totals from parcel and shoelace tasks only.*

Task	Knots & Chiralities	Tied by Sinistrals	Tied by Dextrals
Parcel knot	S Half Knots	48	469
		42% of LH tying, 9% of S knots	65.7% of RH tying, 91% of S knots
	Z Half Knots	66	245
		58% of LH tying, 21% of Z knots	34.3% of RH tying, 79% of Z knots
Shoelace knot	S Half Knots	52	475
		46% of LH tying, 10% of S knots	66.5% of RH tying, 90% of S knots
	Z Half Knots	62	239
		54% of LH tying, 20.6% of Z knots	33.5% of RH tying, 79.4% of Z knots
Totals for shoelace and parcel tasks	S Half Knots	100	944
		44% of LH tying, 9.6% of S knots	66% of RH tying, 90.4% of S knots
	Z Half Knots	128	484
		56% of LH tying, 21% of Z knots	34% of RH tying, 79% of Z knots

TABLE 4—*Numbers of S and Z Half Knots tied first or second during the parcel and shoelace tasks.*

	Hitch	Tied by Sinistrals		Tied by Dextrals	
Task	Chirality	1st half knot (%)	2nd half knot (%)	1st half knot (%)	2nd half knot (%)
Parcel	S	22 (38.6)	26 (45.6)	243 (68.0)	226 (63.3)
	Z	35 (61.4)	31 (54.4)	114 (32.0)	131 (36.7)
Shoelace	S	28 (49.0)	24 (42.0)	251 (70.3)	224 (62.7)
	Z	29 (51.0)	33 (58.0)	106 (29.7)	133 (37.3)
Totals	S	50 (44.0)	50 (44.0)	494 (69.2)	450 (63.0)
	Z	64 (56.0)	64 (56.0)	220 (30.9)	264 (37.0)
Total Half Knots combined	S	100 tied by LH (44.0%)		944 tied by RH (66.0%)	
	Z	128 tied by LH (56.0%)		484 tied by RH (34.0%)	

TABLE 5—*Overhand Knot and Half Hitch totals from the other three tying tasks.*

Task	Knots & Chiralities	Tied by Sinistrals	Tied by Dextrals
3-cord, 3-knot task	S Overhands & Hitches	47 67% of LH tying, 10% of S knots	419 78% of RH tying, 90% of S knots
	Z Overhands & Hitches	23 33% of LH tying, 16% of Z knots	120 22% of RH tying, 84% of Z knots
1-cord, 3-knot task	S Overhands & Hitches	117 62% of LH tying, 11% of S knots	931 74% of RH tying, 80% of S knots
	Z Overhands & Hitches	71 38% of LH tying, 18% of Z knots	325 26% of RH tying, 82% of Z knots
Looped-cord, 4-knot task	S Overhands & Hitches	67 63% of LH tying, 12% of S knots	485 73% of RH tying, 88% of S knots
	Z Overhands & Hitches	40 37% of LH tying, 18.5% of Z knots	176 27% of RH tying, 81.5% of Z knots
Totals for all three tasks	S Overhands & Hitches	231 63% of LH tying, 11% of S knots	1,835 75% of RH tying, 89% of S knots
	Z Overhands & Hitches	134 37% of LH tying, 18% of Z knots	621 25% of RH tying, 82% of Z knots

TABLE 6—*Range of S and Z tying habits according to participant totals.*

Tying Habit	Number of Sinistrals (76)		Number of Dextrals (459)	
S	9 (12%)	S and Sz combined	95 (21%)	S and Sz combined
Sz	29 (38%)	38 (50%)	245 (53%)	340 (74%)
sz	3 (4%)		10 (2%)	
sZ	25 (33%)	Z and sZ combined	99 (22%)	Z and sZ combined
Z	10 (13%)	35 (46%)	10 (2%)	109 (24%)

S, Tied all S; Sz, Tied mostly S; sz, No dominance (tied S and Z equally); sZ, Tied mostly Z; Z, Tied all Z.

perform one task with the left hand other than writing, still produced mostly S enantiomers but not as frequently. The inclination to tie Z knots increased slightly for them. Once again, the probability is less than 0.001 that this data should be evenly distributed.

Claimed Tying Skill Versus Demonstrated Habit: Reef and Granny Tying Habits

When participants were asked to demonstrate how they would tie their shoelaces and secure a parcel or package, most produced Granny Knots or Reef Knots, which consist of two Half Knots. Some participants secured the ends of the cord with more than two Half Knots, and a small number produced anomalies like slipknot variants of the Figure Eight or Crossing Hitch (Fig. 21). These were not included in the tabulated data.

As summarized in Table 8, with a probability of less than 0.001 that the data ratios should be 50%, about one-third of all respondents who tied parcel knots produced Reefs. It is noteworthy that 102 of those claiming Reef knowledge produced a Granny Knot or something else—slightly more than one out of every two participants. Conversely, of the 303 respondents claiming no prior knowledge of the Reef Knot, 50 produced parceling Reef Knots—about one in six.

Table 9 displays detailed data for respondents who tied Reef and Granny Knots during the parcel and shoelace tasks. Overall, most of these respondents tied consistent Grannies for both tasks,

TABLE 7—*Numbers of participants indicating the gradation of handedness relative to tying habit (105 respondents from the original long survey).*

	Tying Habit				
Handedness	S	Sz	sz	sZ	Z
Right-handed (37)	26 (70%)	2 (5.50%)		5 (13.50%)	4 (11%)
Mainly right-handed (60)	30 (50%)	10 (17%)		7 (11.50%)	13 (21.50%)
Ambidextrous (4)	1		2		1
Mainly left-handed (2)	1				1
Left-handed (1)	1				
Totals (105)	59	12	2	12	19

S, Tied all S; Sz, Tied mostly S; sz, No dominance (tied S and Z equally); sZ, Tied mostly Z; Z, Tied all Z.

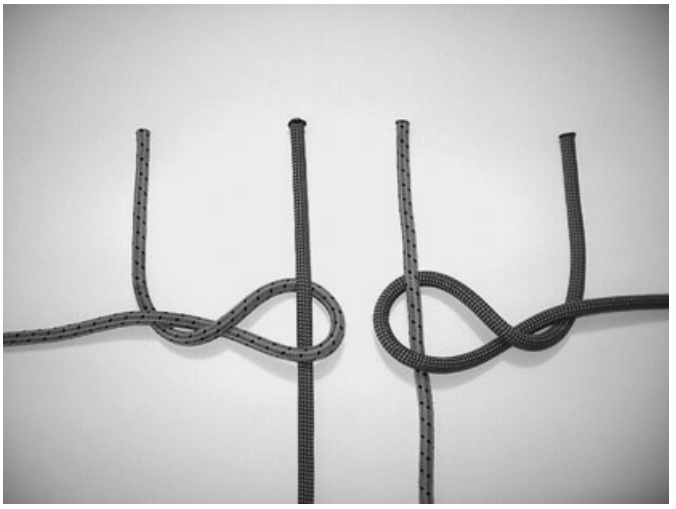


FIG. 21—*Figure eight or crossing hitches, S (left) and Z (right).*

whereas the fewest tiers produced contrary Granny Knots. Of all the respondents who tied Reefs for both tasks, the largest cohorts were consistent Reef Knot tiers, those who tied S/Z or Z/S Reefs for both the parcel and shoelaces tasks. Respondents who tied opposite Reefs (contrary tiers) occurred in the smaller survey cohorts.

Irrespective of handedness, this suggests that the majority of tiers (53% of those tallied) tend to be fairly consistent in their tying habits, whatever the task. Nevertheless, it must be noted that 9% of all respondents tallied were contrary tiers, those who reversed their habit on both Half Knots when they switched from parcel to shoelace tying. The exact mechanism for this reversal is not known, but the only difference between the tasks was the tying of bows or slip loops during the shoelace task. Additionally, 38% of all respondents tallied in Table 9 were inconsistent tiers. This is a substantial percentage. What is the mechanism for the reversal of either Half Knot from task to task, as shown here for inconsistent tiers? Again, the tying of slip loops or bows appears to be the obvious complicating factor.

Table 10 presents this data in a ranked format according to numbers of participants. More dextrals produced S/S instead of Z/Z Grannies, whether tying parcel or shoelace knots, with S/Z being the more common Reef Knot. Surveyed sinistrals tied Z/Z Grannies for both tasks more often than S/S Grannies, and the production of one version of the Reef more commonly than the other was not apparent. The chi-squared test for homogeneity indicates that the probability of equal data ratios is less than 0.001. The Spearman rank correlation coefficient is +0.8 for the parcel knot rankings (strong correlation) and +0.6 for the shoelace knot order (moderate correlation).

Instead of numbers of tiers, individual Reef and Granny Knot counts are presented in Table 11. Dextral tiers produced the S/Z Reef more often than the Z/S version. Sinistrals tied Z/S Reefs slightly more frequently than S/Z Reefs. Once again the data show that dextrals tied substantially more S/S than Z/Z Granny Knots, whereas sinistrals tied more Z/Z than S/S Grannies.

Shoelace Versus Parcel Tying

Of more than 500 participants, 414 completed both the parcel-tying and shoelace-tying tasks. Other participants did not complete these tasks or failed to follow survey instructions, or they tied anomalies or specialized knots like the Bowline, which could not be included in this segment of the analysis.

As shown in Table 12, respondents did not tie parcel knots that always matched their shoelace knots. Some did, but many tied Reef parcel knots and Granny shoelace knots, and *vice versa*. Some even tied knots with opposite Half Knots. Therefore, as a general rule, shoelaces may not be good indicators of nonshoelace tying. All of the useable data are presented according to handedness and tying habits. Ignoring rare tying anomalies, the discrete sample space consisted of 16 possible outcomes, taking into account every permutation and combination of the four Half Knots required to produce two Reef Knots, two Granny Knots, or one of each. Hence, the data are arranged into 16 cohorts. The greatest dextral cohort included tiers of S/S Grannies for both shoelace and parcel tasks. Similarly, the largest sinistral total occurred in the Z/Z parcel and shoelace Grannies cohort. The lowest numbers occurred in cohorts

TABLE 8—*Numbers of participants tying parcel Reef Knots relative to numbers of participants claiming Reef knowledge.*

	Number of Tiers Who Claimed Reef Knowledge (225)	Number of Tiers Who Claimed No Reef Knowledge (303)
Tiers who tied Reef (173)	123 (55% of tiers claiming knowledge) (71% of tiers who tied Reef)	50 (16.5% of tiers claiming no knowledge) (29% of tiers who tied Reef)
Tiers who did not tie Reef (355)	102 (45% of tiers claiming knowledge) (29% of tiers not tying Reef)	253 (83.5% of tiers claiming no knowledge) (71% of tiers not tying Reef)

TABLE 9—*Numbers of participants tying parcel and shoelace Reef and Granny Knots.*

Reef and Granny Tying Habits	Parcel, Shoelace	Number of Sinistral Tiers	Number of Dextral Tiers	Knot Pair Totals	Habit Group Totals
Consistent Reefs	S/Z, S/Z	5	32	37	64
	Z/S, Z/S	4	23	27	
Contrary Reefs	S/Z, Z/S	0	11	11	23
	Z/S, S/Z	3	9	12	
Inconsistent (Parcel Reef, Shoelace Granny)	S/Z, S/S	0	29	29	74
	S/Z, Z/Z	2	8	10	
	Z/S, S/S	1	21	22	
	Z/S, Z/Z	3	10	13	
Inconsistent (Shoelace Reef, Parcel Granny)	S/S, S/Z	4	42	46	84
	Z/Z, S/Z	1	5	6	
	S/S, Z/S	0	9	9	
	Z/Z, Z/Z	5	18	23	
Consistent Grannies	S/S, S/S	11	107	118	155
	Z/Z, Z/Z	15	22	37	
Contrary Grannies	Z/Z, S/S	3	6	9	14
	S/S, Z/Z	0	5	5	

Correction added after online publication 10 May 2010: Column 4, line 5 incorrectly listed a value of "27." It has been corrected to "29."

TABLE 10—*Numbers of tiers ranked according to types of knots tied for parcel and shoelace tasks.*

Number of Sinistral Tiers (57)		Number of Dextral Tiers (357)	
Ranking of parcel knots opposite to dextral order		Ranked parcel knots in descending order	
Z/Z	24 (42.1%)	163 (45.6%)	S/S
Z/S	11 (19.3%)	80 (22.4%)	S/Z
S/Z	7 (12.3%)	63 (17.7%)	Z/S
S/S	15 (26.3%)	51 (14.3%)	Z/Z
Ranking of shoelace knots opposite to dextral order		Ranked shoelace knots in descending order	
Z/Z	20 (35.0%)	163 (45.7%)	S/S
Z/S	9 (16.0%)	88 (24.7%)	S/Z
S/Z	13 (23.0%)	61 (17.0%)	Z/S
S/S	15 (26.0%)	45 (12.6%)	Z/Z

TABLE 11—*Reef and Granny Knot counts according to tier handedness.*

Sources	Reef Knots	Tied by Sinistrals (%)	Tied by Dextrals (%)
Parcel knots	S/Z	7 (39.0)	80 (56.0)
	Z/S	11 (61.0)	63 (44.0)
Shoelace knots	S/Z	13 (59.0)	88 (59.0)
	Z/S	9 (41.0)	61 (41.0)
Total knots, both tasks	S/Z	18 (45.0)	168 (57.5)
	Z/S	22 (55.0)	124 (42.5)
Sources	Granny Knots	Tied by Sinistrals (%)	Tied by Dextrals (%)
Parcel knots	S/S	15 (38.5)	163 (76.0)
	Z/Z	24 (61.5)	51 (24.0)
Shoelace knots	S/S	15 (43.0)	163 (78.4)
	Z/Z	20 (57.0)	45 (21.6)
Total knots, both tasks	S/S	30 (40.5)	326 (77.3)
	Z/Z	44 (59.5)	96 (22.7)

with contrary and inconsistent tying habits, specifically respondents who produced both Z/Z and S/S Grannies, or tiers who produced Z/Z and S/Z or S/S and Z/S combinations. These numbers form an intriguing, nonrandom pattern that deserves further investigation.

The Spearman rank correlation coefficient was +0.787 for this data pattern, indicating a strong correlation, and the Wilcox Signed Rank Test indicated that this rank order is not random.

TABLE 12—*Participant totals ranked according to parcel and shoelace tying cohorts.*

Sinistral Respondents			Dextral respondents		
Parcel Knot	Shoelace Knot	Number of Tiers	Number of Tiers	Parcel Knot	Shoelace Knot
Z/Z	Z/Z	15	107	S/S	S/S
Z/Z	Z/S	5	42	S/S	S/Z
Z/S	Z/S	4	32	S/Z	S/Z
Z/S	Z/Z	3	29	S/Z	S/S
S/Z	S/Z	5	23	Z/S	Z/S
S/S	S/S	11	22	Z/Z	Z/Z
S/Z	Z/Z	2	21	Z/S	S/S
S/S	S/Z	4	18	Z/Z	Z/S
Z/S	S/Z	3	11	S/Z	Z/S
S/Z	S/S	0	10	Z/S	Z/Z
Z/Z	S/Z	1	9	S/S	Z/S
S/Z	Z/S	0	9	Z/S	S/Z
Z/S	S/S	1	8	S/Z	Z/Z
S/S	Z/Z	0	6	Z/Z	S/S
Z/Z	S/S	3	5	S/S	Z/Z
S/S	Z/S	0	5	Z/Z	S/Z

Discussion

Academic research into knot-tying has included investigations focusing on developmental psychology, learning and epistemology (13–16), transitions between graphic and verbal encoding (16), and the anthropological significance of recurring structures (17). There have been a few studies that address tier habit and handedness, but they were comparatively informal investigations conducted by forensic investigators and individuals with practical interests in knots (1–4,18,19).

Handedness and Knot Chirality

The results of one survey suggested there was no correlation between handedness and Granny Knot chirality (4). Nute (2) reported that there may be some link between tier handedness and the twist in the hitches people habitually tie, but his data run contrary to results reported here and by others. Pawson (19) conducted a brief study in which dextrals tended to tie S-twisting Overhand Knots. There was an insufficient number of sinistrals to make any conclusions about their dominant behavior. In 1985, I undertook a

preliminary investigation of knot-tying habits. Dextrals were found to tie S hitches more often, which was consistent with Pawson (19). However, the disagreement with Nute's findings demonstrated that a more rigorous investigation was required. Recently, Spörri (20) reported findings from a small sample survey conducted in Switzerland that also supports Pawson and the results outlined herein.

No matter how the data in this study are analyzed, it is obvious that dextrals produce S more often than Z knots, whether they are using one wend or two wends. In the case of sinistrals, this is not so clear. Although there is a tendency for sinistrals to tie Z Half Knots with two working ends, the data for one-wend tasks suggest sinistrals tie S hitches slightly more often than Z hitches. The mechanism for this is not known but these data are consistent with behavioral research that indicates sinistrals are less consistently left-handed than dextrals are right-handed (21–27).

Right-handed behavior appears to be more global, whereas nearly all left-handers exhibit distinct profiles (28–31). Hand preference in tool use and object manipulation is strongly lateralized in self-professed dextrals and sinistrals. When performing an unskilled motor task, sinistrals have shown greater readiness to use the non-preferred hand than dextrals (39). Left-handers are less consistently sinistral in their manual preference than right-handers are dextral, and many characterizing themselves as left-handed will in fact preferentially use the right hand for a number of activities (29–34). Also, sinistrals are less consistent in their reporting of hand use than dextrals, and the within-subject consistency in hand use for particular tasks among sinistrals is also lower (22–26,33–37).

Even though a relationship between handedness and hitch chirality is apparent, it appears that it is not consistent enough by itself to make exact assessments regarding the handedness of the tier for any given knot sample found at a crime or death scene. Depending on the peculiarities of the case, to do so would likely require the number of potential tiers or suspects to be low and any opinion would benefit from the guidance of additional case information. Also, attention must be paid to the context of the knots, the number of wends employed to create those knots, and situational factors that could cause hitch reversal.

Situational Factors

Although there is a natural tendency for tiers to create primarily S or Z hitches, which could be attributable to the mechanics of the task and tier handedness, situational factors could alter a tier's fundamental habit (1–3).

Nute (2) found that survey participants exhibited consistent tying behaviors over several years, and most could not produce knots of opposite twist. This finding agrees with what is considered common knowledge among knot aficionados (1–4). However, he also noted that this force of habit in tying hitches of the same twist can be over-ridden when individuals tie knots in an awkward position, like behind their heads. Underwood, while investigating multiple murders at Farham in England in the late 1970s, discovered this same reversal phenomenon (3,4).

It was observed in this study that the number of working ends can be another potential hitch-reversal factor. Some tiers tended to naturally tie S hitches with one wend but Z hitches with two wends, and they found it difficult, if not impossible, to reverse this habit (i.e., to tie Z hitches with one wend and S with two). The tying of loops is another factor, as in the case of shoelace tying. For example, some subjects who strongly preferred the S hitch (whether using one or two wends) tied Z hitches when making slip loops or bows in shoelaces.

The issue of how one-wend and two-wend tyings are related to tier handedness was not adequately addressed in other investigations (1–4,18,19). These studies also did not assess the correlation between parcel tying and shoelace tying, the latter of which involves bows or slip loops. The tying tasks and methods of surveying tier handedness were not standardized, and the treatment of tier handedness and habitual tying habits as discrete binary phenomena was overly simplistic.

It was demonstrated that, like handedness, tying habit is not necessarily binary. Those with a lower degree of hand dominance can exhibit hitch-reversal behavior more frequently. How closely the scales of hand preference and tying tendencies correlate is a matter for future inquiry. However, such an inquiry could benefit behavioral researchers but may not assist forensic investigators when it comes to actual case work. This is because measuring hand dominance in any detail, let alone tying habits, is problematic. Controlled conditions and extensive questionnaires are needed for accuracy but these are impractical to administer during criminal investigations.

The periodic occurrence of reverse chirality in a strongly lateralized knot tier's habit could be because of the various hitch-reversal factors mentioned. More research is required to determine how susceptible tiers may be to any one of these influences. The effect Reef Knot training might have on habitual behavior should be considered as well.

Learning and Tying Habits

The results suggest that prior knowledge of the Reef Knot, or lack thereof, does not necessarily translate into a regular tying habit. Learning how to tie a Reef Knot potentially equips tiers with the skill to produce both S and Z Half Knots when using two wends. According to Pawson's informal study (19), the dextrals sampled produced the S/Z version of the Reef Knot more often than its mirror image, which agrees with the reported data. However, his analysis did not address latent habits and previous learning and experience, as was done in this survey.

With regard to claimed Reef knowledge versus displayed tying habits, the results clearly demonstrated that the occurrence of Reefs or Grannies alone is not necessarily a reliable indication of tying sophistication or ignorance. Even respondents who claimed to know and tie the Reef Knot did not use it in practice. Conversely, many unschooled tiers produced Reefs.

Years have been devoted to observing thousands of people tie knots in a variety of situations. These observations and a detailed analysis of more than 100 criminal investigations involving knotted ligatures support the contention that most people in the general population usually tie very simple, unsophisticated knots—mainly Grannies (1–4). This study's presented survey data agree.

Shoelace and Parcel Tying Pattern

The Table 12 ranking pattern illustrates that parcel and shoelace tying habits are phenomena of degrees, which is consistent with handedness research (22,25,27,29–31). The dominance of S versus the Z is linked to tier handedness in a somewhat symmetrical way, and there is a much more complex and intricate relationship at work here that requires further examination. Note that there is a spike in the S/S, S/S cohort for sinistrals (11 respondents), suggesting a tendency for sinistrals to exhibit right-handed habits. This agrees with other handedness research (29–34). Future study could benefit from an extensive pattern analysis that includes handedness gradation rather than a simple division between

dextral and sinistral tiers. However, management of such data would be complex.

Defining and Measuring Handedness

The literature indicates that a blanket attribution of one hand being superior over another in any individual is an oversimplification (27–35). Handedness is a very complex phenomenon. Researchers face two fundamental questions. How should handedness be defined? How can it be reliably measured? Hand preference and skill must be distinguished, and the relationship of the right hand to the left depends on how the various digits are co-ordinated to execute different tasks in a finely programmed sequence of movements, and whether the actions are ballistic or corrective (21,26,29,37–44).

Researchers have relied on four basic methods for the determination of handedness in test subjects: self-reporting, writing hand preference, observation of unimanual activity, and questionnaires (27,28,32–43). Numerous techniques—such as the Edinburgh Handedness Inventory, the Briggs-Nebes Inventory, the Purdue Pegboard test of manual dexterity, and multivariate analysis—have been utilized in research (22–43). The notion of continuously distributed handedness is now widely accepted. Direct observation and questionnaires, which were employed in this study, are currently considered the best measures. Some elements of the aforementioned survey techniques were incorporated. Participant handedness was determined by reported writing hand and dominant hand, and it was verified by responses to questions regarding other handedness tasks.

How have left-handedness, right-handedness, and ambidexterity been defined and interpreted on questionnaires? These decisions have been arbitrary and this accounts for many of the contradictions throughout the handedness literature (21,33), especially prior to the year 2000. People have been classified into three basic types: dextral, sinistral, and ambidextral—or, more recently, mixed, inconsistent, or shared handedness (22,24–26,33–35). It is now clear that this simple classification does not account for the full range of manual asymmetry that can be observed.

Of any asymmetrical activity undertaken in our society today, handwriting is used most frequently to distinguish between sinistrals and dextrals (45,46). More sophisticated surveys may be designed in the future to account for the full range of lateralization relative to tying habits (20). However useful it may be from a research perspective, this approach will likely have little if any application to actual case work. Writing hand and reported dominant hand will continue to be the primary indicators because they are the simplest to assess.

Strengths and Limitations of this Study

This survey was superior to previous work in several ways. First, the total number of participants surveyed was large—more than the sum of all respondents from all other studies combined. Second, the survey tying tasks were standardized, and each person had the opportunity to perform and even repeat several different tying tasks.

Nevertheless, several limitations must be noted. Like every other statistical investigation of knot-tying behavior, this survey is based on a sample that is not perfectly random, although an effort was made to survey a number of different demographic groups. Considering time and budget constraints, not to mention the ethics of enlisting voluntary participants, this ideal cannot be achieved easily.

It would be interesting to sample an inmate population for comparison as data acquired from those subjects could be more applicable to criminal case investigations. However, legal and ethical issues concerning anonymity, confidentiality, and voluntary participation would complicate data collection.

According to previous research, the frequency of true left-handers in the general population is one in 14 to one in 10 people (1,3). The occurrence of sinistrals in this survey was higher: approximately one in six respondents. Even still, that makes it difficult to acquire sufficient comparative data pertaining to left-handers, and the reported data suffered in that regard. In a recent knot-tying study (20), an even number of right-handers and left-handers (10 in each cohort) were categorized using the Edinburgh inventory.

Those who exhibit mixed, shared, or ambidextral handedness are even rarer. Just four participants in this study indicated shared, inconsistent, or ambidextral tendencies—about one in every 136 survey participants—and the incomplete data for those individuals were not included.

Also, many survey respondents were knowledgeable, experienced knot tiers. This experience may have obscured some of the latent habits respondents might have displayed as children. However, as the data indicated, claimed Reef knowledge or ignorance was not always consistent with tying habits.

The other tasks—single cord tyings, forced one-wend tyings, and so forth—were included in an effort to determine natural tying habits relative to handedness. Survey respondents could not be monitored individually all of the time as they completed these tasks. Hence, misinterpretations of instructions occurred and some surveys were left unfinished. It would be ideal to observe each subject separately and digitally record their actions for later analysis.

Conclusions

The results of this study did not indicate an exact correlation between the principal manipulating hand and the chirality of resultant knots. Overall, right-handed subjects tended to tie S-twisting hitches more often than Z-twisting hitches. Sinistrals demonstrated S tying less frequently and even tied Z knots more often in some situations.

Not everyone claiming to know the Reef Knot actually used it to secure the shoelace cord, parcel cords, or both. Additionally, of those producing Reef Knots, only about one-third employed the knot to secure both the shoelace and parcel cords. This suggests that any acquired knowledge of the Reef Knot may not be manifested in regular tying habits. A number of individuals claiming to have no knotting experience or training produced Reef Knots. This indicates that some may learn the Reef Knot through personal discovery, or tie it out of habit.

Those participants producing Reef Knots tied either the S/Z or the Z/S version. Hand preference did not seem to be strongly correlated with the type of Reef Knot tied, although S/Z Grannies were more abundant than Z/S among dextrals. As well, parcel and shoelace knots were not always consistent, although an interesting and complex mirror-image gradation pattern was revealed that was symmetrical according to writing hand.

The number of available working ends may affect what a tier produces. The use of one working end could produce hitches of one particular type, but the utilization of two working ends can result in hitch reversal for some tiers. Therefore, likely one-wend and two-wend tyings should be recognized in case evidence.

These results serve as a foundation for future research in forensic knot analysis. Obtaining more information about the latent tying habits of inexperienced knotters would be valuable. Parceling knot

data are probably the most applicable to case work, because tying parcels is similar to securing neck and limb ligatures.

Self-tying needs to be studied further under safely controlled conditions. It would be useful to perform a qualitative study of knot tying habits and make a video record of each subject's knotting activity, especially if it could be coupled with a qualitative investigation of how people learn and remember knots, and with an examination of the process by which they recall and apply knotting knowledge. Another potentially productive avenue would require a greater number of subjects from the general population and a revised quantitative survey, one that measures the range of preferred handedness and chirality more thoroughly, although writing hand will probably remain the key determinant in case work.

Acknowledgments

I express my gratitude to Inspector (retired) Geoffrey Budworth, formerly of the Metropolitan Police, London, England and founding member of the International Guild of Knot Tyers for inspiring my interest in forensic knot analysis and for his long-distance feedback and mentorship over the past three decades. My former graduate studies supervisor, Professor Emeritus Robert Horwood, deserves thanks for overseeing my initial survey work that eventually evolved into this bigger research project. Thanks are due to Mr. Des Pawson MBE, also a founding member of the International Guild of Knot Tyers, for collecting data in Great Britain. I am indebted to several friends, including Sandra Berg of The Royal Military College of Canada and Hai Pham of Agriculture, Canada (Bioinformatics) for reading early drafts, and especially Dr. Ellen Tsai of Queen's University for providing invaluable editorial feedback. Thank you all.

References

1. Chisnall R. What knots can reveal: the strengths and limitations of forensic knot analysis. J Forensic Ident 2007;57:726–49.
2. Nute HD. Mirror images in knots. J Forensic Sci 1986;31:272–9.
3. Chisnall R. The forensic analysis of knots and ligatures. Salem, OR: Lightning Powder Company, Inc, 2000.
4. Budworth G. Knots and crime. Great Britain: Police Review Publishing Co Ltd, 1985.
5. Budworth G. Identification of knots. J For Sci Soc 1982;22:327–31.
6. Hazelwood R, Dietz AE, Berges AW. Autoerotic fatalities. Toronto: Lexicon Books, 1981.
7. Ashley CW. The Ashley book of knots. Garden City, NY: Doubleday & Company, Inc, 1944.
8. Sumners DW. The role of knot theory in DNA research. Geometry and topology: manifolds, varieties, and knots. In: McCrory C, Shifrin T, editors. Lecture Notes in Pure and Applied Mathematics. New York: Marcel Dekker, Inc., 1987;Vol. 105:297–318.
9. Van de Griend P. A history of topological knot theory. In: Turner JC, van de Griend P, editors. The history and science of knots. New Jersey: World Scientific, 1996;205–60.
10. Canadian General Standards Board. Standards CAN2-4.2-M77, Method 9.4, 1984.
11. Canadian Government Specifications Board. Standards 40-GP-1M, 1978.
12. Wendrick W. Ancient Egyptian rope and knots. In: Turner JC, van de Griend P, editors. The history and science of knots. New Jersey: World Scientific, 1996;43–68.
13. Gruber HE, Vonèche JJ, editors. The essential Piaget. New York: Basic Books, 1977.
14. Olson DR. Cognitive development: the child's acquisition of diagonality. New York: Academic Press, 1970.
15. Eliot J, Salking NJ, editors. Children's spatial development. Springfield, IL: Charles C. Thomas Pub Ltd, 1975.
16. Strohecker C. Why knot? [dissertation]. Cambridge, MA: Massachusetts Institute of Technology, 1991.
17. Turner JC, van de Griend P, editors. History and science of knots. New Jersey: World Scientific, 1996.
18. Chisnall R. A survey of tier handedness and habit—some preliminary data. Knotting Matters Early 1994;45:9–14.
19. Pawson D. Left hand—right hand? Knotting Matters, Late 1992;41:14.
20. Spörri S. La valeur indicale des noeuds [dissertation]. Lausanne-Dorigny, Switzerland: University of Lausanne, Ecole des sciences criminelles, Institut de Police Scientifique, 2008.
21. Kimura D, Vanderwolf CH. The relation between hand preference and the performance of individual finger movements by left and right hands. Brain 1970;93:769–74.
22. Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 1971;9:97–113.
23. Annett J, Annett M, Hudson PTW, Turner A. The control of movements in the preferred and non-preferred hands. Quart J Exper Psychol 1979;3:641–52.
24. Annett M. Five tests of hand skill. Cortex 1992;28:583–600.
25. Dragovic M, Hammond G. A classification of handedness using the Annett Hand Preference Questionnaire. Br J Psychol 2007;98(3):375–87.
26. Todor JI, Doane T. Handedness classification: preference versus proficiency. Percept Mot Skills 1977;45:1041–2.
27. Corey DM, Hurley MM, Foundas AL. Right and left handedness defined: a multivariate approach using hand preference and hand performance measures. Neuropsychiat Neuropsychol Behav Neurol 2001; 14(3):144–52.
28. Doyen AL, Duquenne V, Nuques S, Carlier M. What can be learned from a lattice analysis of [Annett's] laterality questionnaire? Behav Genet 2001;31(2):193–207.
29. Steenhuis RE, Bryden MP. Different dimensions of hand preference that relate to skilled and unskilled activities. Cortex 1989;25(2):289–304.
30. Healey JM, Liederman J, Geschwind N. Handedness is not a unidimensional trait. Cortex 1986;22(1):33–53.
31. Calvert GA, Bishop DVM. Quantifying hand preference using a behavioural continuum. Laterality 1998;3 (3):255–68.
32. Bryden MP. Measuring handedness with questionnaires. Neuropsychologia 1977;15:617–24.
33. Annett M. The binomial distribution of right, mixed and left handers. Quart J Exper Psychol 1967;19:327–34.
34. Annett MA. Classification of hand preference by association analysis. Br J Psychol 1970;61(3):303–21.
35. Ransil BJ, Schachter SC. Test-retest reliability of the Edinburgh Handedness Inventory and global handedness preference measurements, and their correlation. Percept Mot Skills 1994;79(3 Pt 1):1355–72.
36. Tracy J, Flanders A, Madi S, Laskas J, Stoddard E, Pyrros A, et al. Regional brain activation associated with different performance patterns during learning of complex motor skills. Cereb Cortex 2003;13(9):904–10.
37. Flowers K. Handedness and controlled movements. Br J Psychol 1975;6:39–52.
38. Steingruber HJ. Handedness as a function of test complexity. Percept Mot Skills 1975;40:263–6.
39. Provins KA. Motor skills, handedness and behaviour. Aust J Psychol 1967;19:137–50.
40. Miller E. Handedness and the pattern of human ability. Br J Psychology 1971;62(1):111–2.
41. Peters M. Handedness effect of prolonged practice on between hand performance differences. Neuropsychologia 1981;19:570–90.
42. Loo R, Schneider R. An evaluation of the Briggs-Nebes modified version of Annett's handedness inventory. Corte 1979;15(4):683–6.
43. Verdino M, Dingman S. Two measures of laterality in handedness: the Edinburgh Handedness Inventory and the Purdue Pegboard test of manual dexterity. Percept Mot Skills 1998;86(20):476–8.
44. Annett M. Handedness and brain symmetry: the right shift theory. East Essex, UK: Psychology Press, 2002.
45. Coren S, Porac C. Normative data on hand position during writing. Cortex 1979;15:679–82.
46. Moscovitch M, Smith LC. Differences in neural organisation between individuals with inverted and non-inverted handwriting postures. Science 1979;205:710–2.

Additional information—reprints not available from author:
Robert C. Chisnall, M.Ed.
60 Dauphin Avenue
Kingston, ON K7K 6B1
Canada
E-mail: chisnall@kingston.net

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01400.x
Available online at: interscience.wiley.com

PAPER

GENERAL

Stephanie K. Bell,[1] *B.S., B.A. and Piero R. Gardinali,*[1,2] *Ph.D.*

Assessment of Silicone Polymer Composites for Environmental Forensic Applications: A Proof of Concept Study*

ABSTRACT: This study evaluates the use of polydimethylsiloxane polymer composites (PDMS, Fe–PDMS) as a passive sampling media to preconcentrate analytes found in environmental settings. Samplers were made using commercially available silicone products. The composite samplers were assessed for their sorption properties using Atrazine and Irgarol 1051 as model compounds. The initial study assessed the utility of PDMS sheets as adsorption material by following analyte depletion from spiked water samples by solid-phase microextraction gas chromatography/mass spectrometry (GC/MS). Follow-up studies conducted at high and low concentrations using lab manufactured iron- PDMS rods (Fe–PDMS) showed effective uptake at differential rates from concentrations ranging between 1 μg/L and 10 μg/L. Adsorption mechanism was reversible, and compounds were recovered from the exposed materials and analyzed by liquid–liquid extraction-GC/MS. Both composites showed better affinity for Irgarol 1051, 100% removal, than for Atrazine, 30% removal, likely representing their K_{OW} differences, 3.6 and 2.6, respectively. This "proof of concept" study demonstrates the positive implications for the use of silicon polymer composites as a monitoring tool for environmental forensic purposes.

KEYWORDS: forensic science, environmental forensics, herbicides, passive sampling, silicone polymer composites, iron-polydimethyl siloxane

Herbicides account for 75% of all pesticides used in the U.S. Vegetation control substances found in aquatic environments originate from both agriculture as well as urban landscapes and are easily transported between compartments via water runoff (1). Therefore, water analysis is the preferred matrix to assess their environmental occurrence. However, water measurements are discrete in nature and not reflective of temporal trends. Passive samplers such as semipermeable membrane devices (SPMDs) are a much more versatile tool to integrate environmental concentrations and to avoid the use of live sentinel organisms for environmental monitoring (1).

Two common herbicides found in aquatic systems include Atrazine and Irgarol, the model compounds for this study. Atrazine is the most commonly used triazine-based pesticide in the U.S., and it has become ubiquitous in the aquatic environment. It is readily found in freshwater environments at concentration between 10 ng/L and 100 ng/L and has a regulatory criterion of 3 μg/L (2–5). Atrazine is extensively used in agricultural areas including corn and related silage crops (4) as well as sugarcane fields (5). It is easily detected because of its ability to persist in soil and its water mobility. Atrazine has been found to produce detrimental environmental effects including toxicity to aquatic plants and insects, phytoplankton, and fish and has recently been identified as a potential endocrine disruptive chemical (2,6–8).

Irgarol is also a triazine-based algaecide but is used to boost the effects of copper-based antifouling paints for boats and vessels. Irgarol is primarily used to inhibit the growth of copper-resistant fouling organisms such as algal slimes and seaweed (9–11). It seeps slowly from the paint, and its presence has been documented in marine environments worldwide (11).

Because triazines occur in low concentrations in aquatic systems, samples must be preconcentrated for trace-level determination. Currently, conventional analyses of triazines present in aquatic systems require sample collection of large volumes of water (1–2 L) to achieve relevant detection limits (low part per trillion, ng/L). It would be ideal to extract triazines and other mildly polar pollutants from smaller volumes or directly from the sampling site. Typical analytical protocols use liquid–liquid extraction, solid-phase extraction, or solid-phase microextraction (SPME). The use of passive samplers is not novel but is often overlooked because of inherited analytical challenges. SPMDs have also been proposed as an "*in-situ* preconcentration" alternative for triazine herbicides (12). The use of SPMDs, however, presents a problem in environmental analysis. The membranes are very hydrophobic, and some of the target compounds remain trapped in the membrane and are difficult to recover (13). This irreversible adsorption problem makes it difficult to accurately determine the mass balance of analytes both in and out of the membrane. In addition, the triolein used as a carrier is often expensive and difficult to clean up, making the process time consuming and cumbersome (1,13).

This study proposes the use and validation of silicon-based polymers, mainly polydimethylsiloxane (PDMS), as preconcentration

[1]Department of Chemistry and Biochemistry, Florida International University, University Park, Miami, FL.
[2]Southeastern Research Center (SERC), Florida International University, University Park, Miami, FL.

*Preliminary information presented at the 57th Annual Meeting of the American Academy of Forensic Sciences, February 21–26, 2005, in New Orleans, LA, and at the SETAC-NA 2005 Annual Meeting, November 13–17, 2005, in Baltimore, MD.

Received 31 Oct. 2008; and in revised form 20 June 2009; accepted 29 June 2009.

devices for common herbicides found in aquatic environments and their evaluation as passive samplers for environmental forensic applications. Important characteristics of silicones include low surface tension, nonionic/nonpolar characteristics, hydrophobicity, thermal stability, oxidation resistance, low toxicity, and most importantly, they are environmentally safe. Silicones offer three good properties to be considered as passive samplers; they have the potential to concentrate analytes (i.e., SPME), they are thermally resistant so they can be properly cured and cleaned, and they resist chemical treatment so they can be solvent extracted for analyte recovery. In addition, silicone polymers can be shaped into forms amenable to on-site deployment. In this regard, as demonstrated in this study, they could also be made as iron-containing composites allowing for easy retrieval through magnetic filtration and recovery (14).

Materials and Methods

Materials

Atrazine (2-chloro-4-ethylamino-6-isopropylamino-*s*-triazine) and TCMX (tetrachloro-m-xylene), used as a recovery standard, were purchased from Ultra Scientific (North Kingstown, RI). Irgarol 1051 (2-methylthio-4-*tert*-butylamino-6-cyclopropylamino-s-triazine) was provided by Ciba Specialty Chemicals (Tarrytown, NY), and deuterated Atrazine (Atrazine d-5) used as a surrogate standard was purchased as a certified standard solution from Dr. Ehrenstorfer GmbH (Augsburg, Germany). Optima™ grade solvents, methylene chloride, and hexane were purchased from Fisher Scientific (Suwannee, GA). The PDMS was purchased from Wal-Mart (Silicon II, General Electric Huntersville, NC). The iron used in the preparation of the Fe–PDMS composites (Iron Powder, approximately 325 mesh, 97%) was purchased from Aldrich Chemical Company, Milwaukee, WI.

Silicone Disks

PDMS disks (1/4″ diameter × 1/8″ thick) were cut from commercially available silicone sheets (McMaster-Carr Atlanta, GA) and used without further manipulation.

Manufacturing of Fe–PDMS Composite Disks/Rods

The Fe–PDMS composites were made using a 1:2 ratio of iron to PDMS. The PDMS was thoroughly mixed with powdered iron until a homogeneous color was achieved (approximately 2 min). The mixture then was extruded into 3/16″ ID silicone tubing (Masterflex, Vernon Hills, IL). The Fe–PDMS was cured, inside the silicone tube, in a drying oven for 4 days at 75°C. The Fe–PDMS material was extracted from the tubing and cut into 1 cm length pieces (0.3–0.5 g each). The rods were then fully cured in a convection oven at 250°C for 2 h.

Composite Cleaning and Activation

Both types of composites (disks and rods, PDMS and Fe–PDMS) were cleaned by solvent extraction using sonication in hexane for 30 min. This extraction process was completed in triplicate. The composites were then dried in a convection oven at 50°C for 1 h and activated at 250°C for at least 4 h before using them for extraction/deployment.

Images of the Fe–PDMS composite surface were obtained using a scanning electron microscope to assess the homogeneity of the material. Figure 1 shows the backscatter electron imaging of a section of a Fe–PDMS rod depicting the conductive iron particles (white) dispersed in the PDMS matrix (gray).

Initial Demonstration of Adsorption Capabilities Study

Fourteen 1-L bottles were filled with 500 mL of deionized water and spiked with 50 μL of 10 ppm (mg/L) of Irgarol 1051 and 50 μL of 10 ppm (mg/L) of Atrazine giving a final target concentration of 1 ppb (μg/L) for each herbicide. All bottles were then placed in a sectioned box attached to an orbital shaker. The box served a twofold purpose: to secure the bottles and also to ensure that photodegradation was not a viable mechanism for analyte removal. The bottles were shaken for 1 min to ensure proper mixing, and 2 mL were collected to calculate the initial concentrations. After the initial collection, 20, 40, and 60 PDMS disks were added to the spiked samples in triplicate. Twenty eight Fe–PDMS rods were also placed in the remaining bottles in triplicate. A blank (negative control) and fortified deionized water (positive control) were also included in the study to monitor system cleanliness and nonabsorptive losses (photo- and or bio-degradation), respectively. The orbital shaker was maintained at enough speed to assure adequate mixing of the sample (100 rpm) throughout the experiment. Water samples (5 mL) were collected every 6 h for the duration of the experiment (80–120 h) for chemical analysis. Before processing, all samples were spiked with 100 μL each of 1 ppm (mg/L) TCMX and 1 ppm (mg/L) Atrazine d-5 (the recovery and surrogate standards, respectively). Samples were analyzed for Atrazine and Irgarol by gas chromatography/mass spectrometry (GC/MS) using a 65 μm PDMS divinylbenzene (PDMS/DVB) SPME fiber (Supelco; Sigma-Aldrich Company, St. Louis, MO). The fiber was initially conditioned on the inlet of the HP5890 Series I Gas Chromatograph at 250°C for 3 h. Each of the collected water samples were extracted with the optimized SPME protocol (30 min, room temperature, full immersion) and analyzed by HP 5890 Series Gas Chromatograph coupled to a HP 5970 mass selective detector operated under electronic impact ionization at 70 eV in selected ion monitoring mode (SIM). The chromatographic separation was carried out in a 30 m × 0.25 μm ID × 0.25 mm film thickness DB-5MS fused capillary column

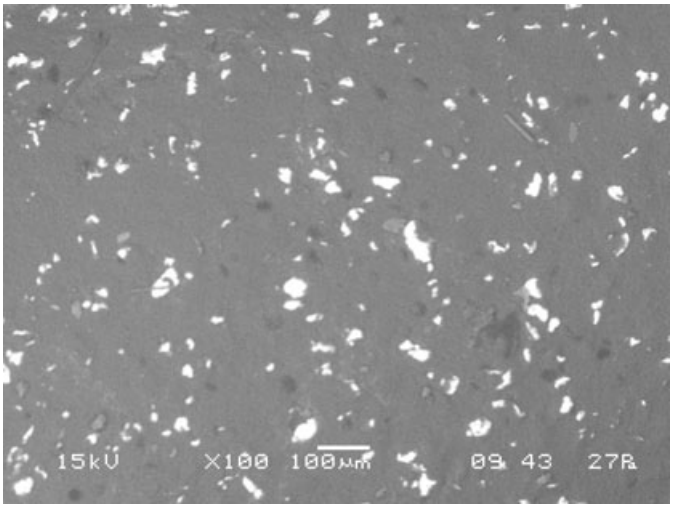


FIG. 1—*SEM image (100×) of the cross-section of a Fe–PDMS rod showing the conductive iron particles (white).*

(J&W Scientific, Folsom, CA). The GC was operated in splitless mode at 250°C. The GC oven was programmed with an initial temperature of 100°C for 1 min and ramped to 300°C at a rate of 15°C/min and held at 300°C for 1 min giving a total run time of 15 min. Helium was the carrier gas at a constant flow of 2 mL/min.

Adsorption Study from Aqueous Samples Using Fe–PDMS Composites

The initial PDMS experiments were repeated to test the Fe–PDMS composites at two different analyte concentrations. High-level (10 μg/L) and low-level (1 μg/L) target concentrations were used to test the material. Five hundred milliliter of DI water was added to each of four 1-L bottles and spiked with both Irgarol 1051 and Atrazine to obtain the final target levels. The bottles were then placed in the box on the orbital shaker and processed for 1 min to ensure the proper mixing of the herbicides. As in the previous step, 5 mL of the initial solutions was collected and analyzed at the beginning of the experiment to assess the initial concentrations. One bottle served as a positive control (degradation check), and one bottle served as the negative control (blank). Fe–PDMS composites were weighed so that approximately equal amounts of the composites were used in each bottle (30–35 pellets). The shaker was operated at 100 rpm, and 5 mL water samples were collected every 6 h for the duration of the experiment (80–120 h). Each water sample was then spiked with 100 μL of 1 ppm (mg/L) Atrazine d-5 (the surrogate standard) and liquid–liquid extracted against 2 mL of methylene chloride in triplicate. The methylene chloride extracts were separated from the water samples using Pasteur pipettes and transferred to Kuderna-Danish (KD) concentrator tubes by passing them through a funnel containing glass wool and anhydrous Na_2SO_4, to remove any remaining water. Teflon boiling chips were added to the KD tubes, and the organic extract was concentrated down to 1 mL using a water bath at 65°C. Each sample was then spiked with 100 μL of 1 ppm (mg/L) TCMX (the recovery standard) and analyzed by GC/MS-SIM. The same parameters were used as previously stated, although the analytical determination was conducted on a Finnigan Trace DSQ GC/MS (Thermo Electron Corporation, San Jose, CA) to improve the detection limits.

Recovery Study from the Fe–PDMS Exposed Pellets

At the end of the adsorption experiments, the Fe–PDMS rods were extracted to assess the uptake of Irgarol 1051 and Atrazine from the water samples. The rods/pellets were removed from the water by using a TFE coated magnetic rod, air-dried for 30 min and placed in 125-mL flasks. Each flask was spiked with 100 μL of 1 ppm (mg/L) Atrazine d-5 (the surrogate standard). Thirty milliliters of methylene chloride was added to each of the flasks, and the rods/pellets were sonicated for 45 min. The solvent was then decanted and filtered through Na_2SO_4 and glass wool and transferred to a second 125-mL round-bottom flask containing Teflon boiling chips. This process was repeated three times. The extracts were then concentrated using a KD tube to a final volume of 1 mL using a water bath at 65°C. Immediately before injection, each sample was spiked with 100 μL of 1 ppm (mg/L) of TCMX (the recovery standard) and analyzed by GC/MS-SIM using the same parameters as before with the Finnigan Trace DSQ GC/MS. Typical chromatograms obtained from the unexposed rods/pellets before and after the uptake experiments are shown in Figs. 2 and 3, respectively.

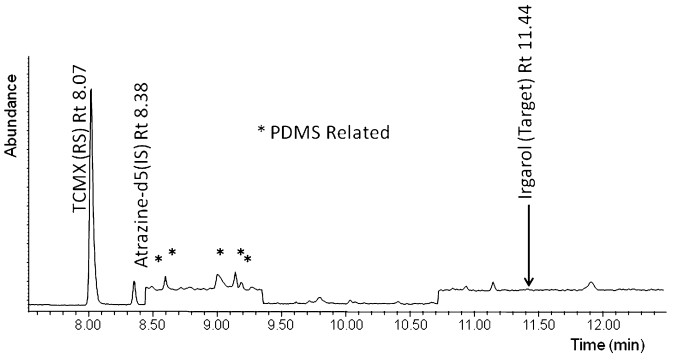


FIG. 2—*Reconstructed mass chromatogram showing the methylene chloride extraction of Fe–PDMS rods exposed to DI water only (BLANK). Peaks in the chromatogram, likely from the silicone material do not interfere with the detection of the analytes in selected ion monitoring mode.*

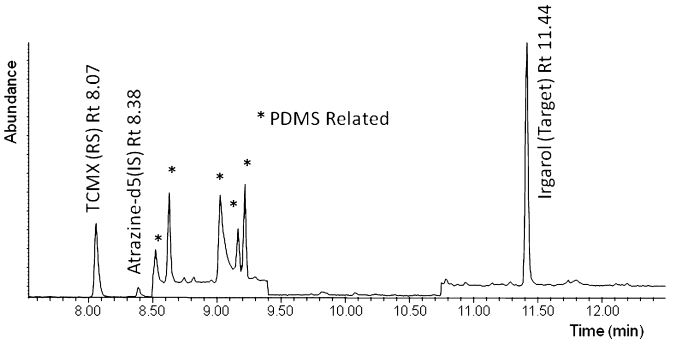


FIG. 3—*Reconstructed mass chromatogram showing the methylene chloride extraction of Fe–PDMS rods exposed to DI water fortified with Irgarol 1051 for 96 h. Atrazine when present is partially resolved from the per-deuterated standard but signals are fully resolved by the mass difference of both compounds.*

Results and Discussion

Initial Adsorption Study (PDMS and Fe–PDMS)

This experiment was used to demonstrate the "proof of concept" of using PDMS composites as passive samplers. A SPME-based method was developed to follow the uptake of Irgarol 1051 and Atrazine by both the PDMS and Fe–PDMS disks and rods. The results of the experiment, presented as the depletion of the analyte in the water samples, are shown in Figs. 4–7 for the two model compounds. Figure 4 shows a rapid uptake of Irgarol 1051 as a function of time with concentrations reaching the method detection limit at about 72 h. As expected, the removal of the herbicide is dependent on the amount of the material while no signs of saturation were observed from the plots. In contrast, Fig. 5 shows the same data for Atrazine. The more pronounced difference is the rate of removal for both herbicides. The calculated first-order half life for Irgarol and Atrazine was 5.54 h and 16.2 h, respectively for 40 PDMS disks. This effect is likely related to the difference in the octanol–water partition coefficient (K_{OW}) for both compounds (Irgarol = 3.6 and Atrazine = 2.6), a likely indicator that the PDMS disks could be used as surrogates for aquatic organisms. Figures 6 and 7 indicate that Fe–PDMS composites are also effective as samplers, but the introduction of the iron limits the availability of the PDMS sites; thus, the adsorption rates are slower. Half lives for Irgarol 1051 and Atrazine were 11.1 h and 20.5 h,

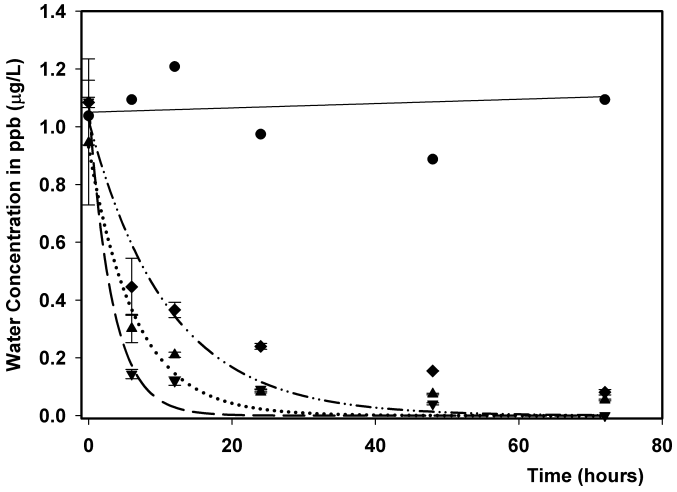


FIG. 4—*The relative concentration of Irgarol 1051 over time is shown for bottles containing 20 discs (♦), 40 discs (▲), and 60 discs (▼). The control (●) shows the constant Irgarol concentration.*

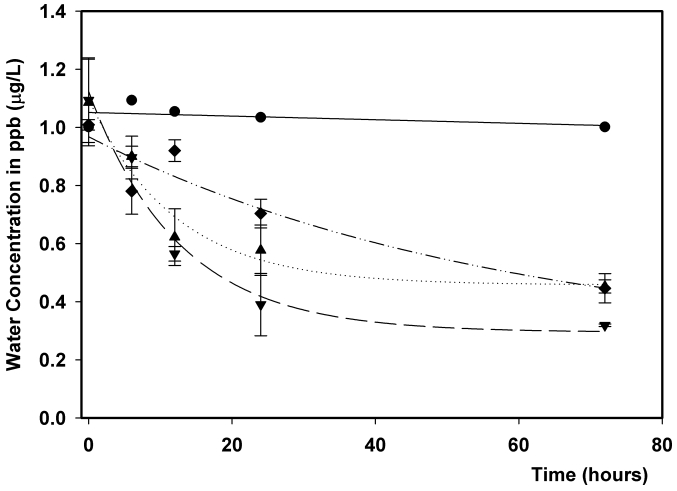


FIG. 5—*The relative concentration of Atrazine over time is shown for bottles containing 20 discs (♦), 40 discs (▲), and 60 discs (▼). The control (●) shows the constant Atrazine concentration.*

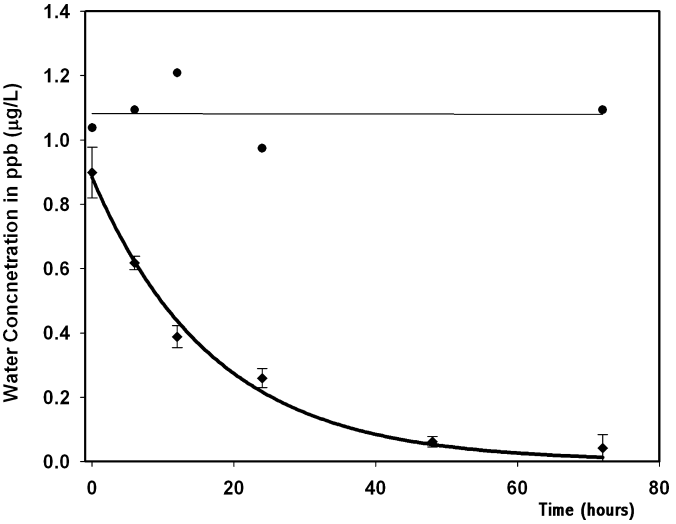


FIG. 6—*The relative depletion of Irgarol 1051 concentration is shown for bottles containing Fe- PDMS (♦). The control bottle (●) shows the constant Irgarol 1051 concentration over time.*

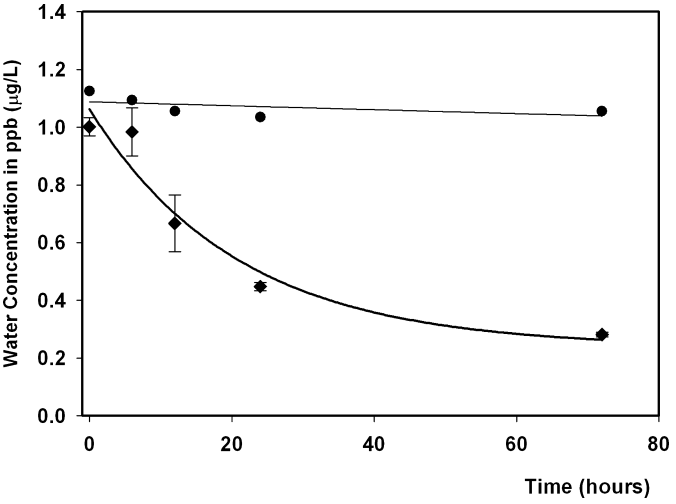


FIG. 7—*The relative depletion of Atrazine (as Atrazine-d5) concentration is shown for bottles containing Fe- PDMS (♦). The control bottle (●) shows the constant Atrazine concentration over time.*

respectively. The Irgarol 1051 adsorption results with the Fe–PDMS indicate approximately 71% depletion in 24 h and 95% depletion of concentration in 72 h. The depletion of the Atrazine was not as efficient with reductions of 55% and 72% in 24 h and 72 h, respectively. This study clearly demonstrated the capabilities of not only PDMS but also Fe–PDMS composites as possible monitoring devices for herbicides in aquatic conditions. The intrinsic advantage of Fe–PDMS is that the material is magnetic. Because of this fact, the pellets could be deployed and recovered from environmental settings with the aid of magnets or easily used for magnetic filtration from contaminated wastewaters.

Fe–PDMS Study: High and Low Concentration

Once the capabilities of the PDMS were established, a similar study was conducted using only Fe–PDMS. In the low concentration study, the same initial concentration of Irgarol 1051, as in the SPME study, was used. Four bottles were spiked to give final concentrations of 1 ppb (1 µg/L). Three of these bottles contained *c*. 6 g of the Fe–PDMS rods each while the fourth bottle was used as a positive control to ensure the analyte did not degrade for the duration of the experiment. In the first 24 h, an average 58.2% of the initial concentration remained detectable in the water samples. After 96 h, the initial 1 ppb (1 µg/L) concentration of Irgarol 1051 was no longer detectable in the water samples (Fig. 8), indicating a 99.9% removal of the analyte by the Fe–PDMS disks. Controls showed relatively constant Irgarol 1051 concentration throughout the experiment, indicating no significant degradation of the analyte in the absence of the composites. The calculated half life for the removal process was 30.53 h. Solvent extraction of the exposed Fe–PDMS disks produced an average of 85.5 ± 1.5% recovery of the Irgarol 1051 based on the amount of herbicide spiked in the water samples. These results clearly indicate the reversible behavior of the Fe–PDMS as adsorption material and the potential for its use as a passive sampling device.

Results were more encouraging for the high concentration study where the potential for material saturation was tested. The

experiments were run in a similar fashion as previously reported, and the results are presented in Figs. 9 and 10. Once again, the consistency of the control samples indicates that there was no degradation of either herbicide during a 3-week period. As in both the preceding studies, the Fe–PDMS removal for Irgarol was much faster than for Atrazine. While 82.9% of Irgarol 1051 was removed in the first 24 h, only 4.5% of the Atrazine was depleted from the water sample. After 96 h, Irgarol was in equilibrium between the water and the rods with an average removal of 87.8% (Fig. 9). Uptake of Atrazine never reached equilibrium before the experiment was stopped. Although concentration depletion for Atrazine was apparent, only 29.7% of the initial concentration was removed by the Fe–PDMS rods (Fig. 10) in the first 96 h. Analysis after 264 h resulted in only 41.9% reduction of the initial 10 μg/L concentrations. The calculated half life for Irgarol 1051 and Atrazine are 23.94 and 185.14 h, respectively.

Solvent extraction of the Fe–PDMS disks yielded an average of 101.7 ± 13.1% of the initial Irgarol 1051 spiked in the water (Fig. 9, right side). The results however were discouraging, but expected for Atrazine. There was an average of 40.3 ± 8.2% recovery of the Atrazine that was deemed adsorbed by the Fe–PDMS composites. It is clear that Atrazine did not have the same rate of uptake as Irgarol 1051 in this study likely reflecting the differences in their partitioning ability. The K_{OW} for Irgarol 1051 is 3.6, while the K_{OW} for Atrazine is 2.6. Although the uptake and recovery of both herbicides differed, this study indicates the success in the use of the Fe–PDMS disks as preconcentration devices for analytes found in aquatic environments.

Conclusions

This study clearly demonstrated the capabilities of silicone-based polymer composites for the extraction of trace herbicides from water samples at environmentally relevant concentrations. Both the PDMS and Fe–PDMS materials studied showed the capacity and robustness to grant their use as passive samplers. Significant reductions in concentrations were observed for both model compounds studied. Irgarol 1051 had a higher rate of adsorption by both the PDMS and Fe–PDMS disks. In addition, an excellent rate of recovery was obtained from exposed rods. Atrazine uptake, however, was not as good as expected. Structural dependence aside, the trend seems to indicate that the ability of PDMS composites to monitor contaminants in aquatic environments is controlled by the same parameters that affect their partition to biotic and abiotic compartments such as K_{OW}. Further studies with more hydrophobic pollutants already underway indicate that Fe–PDMS disks are indeed good integrators of environmental exposure and could, therefore, be used for site deployment. The magnetic properties of Fe–PDMS are also important, because it will allow for easy retrieval from environmental sites and could open the possibility for remediation work. The findings of this study may be particularly important for forensic applications in which confirmation of the presence of a compound, rather than its concentration, is required. As an extension of this work, new experiments are underway to assess the ability of the

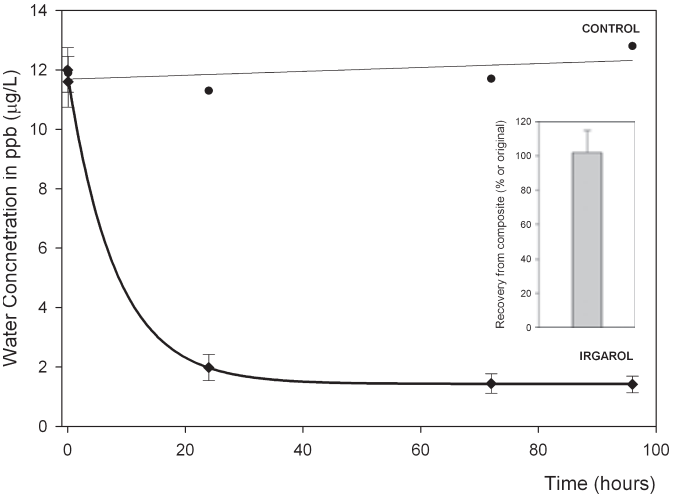


FIG. 9—*Uptake and release of Irgarol 1051 from the Fe- PDMS pellets. Initial water concentration was 10 μg/L (●) and remained constant during the exposure experiment. Uptake curve shows removal of approximately 87.9% of the initial concentration (◆) after 4 days of exposure. Re-extraction of the target compound from the exposed Fe–PDMS pellets produced quantitatively recoveries (102 ± 13%) of Irgarol 1051 based on initial concentrations.*

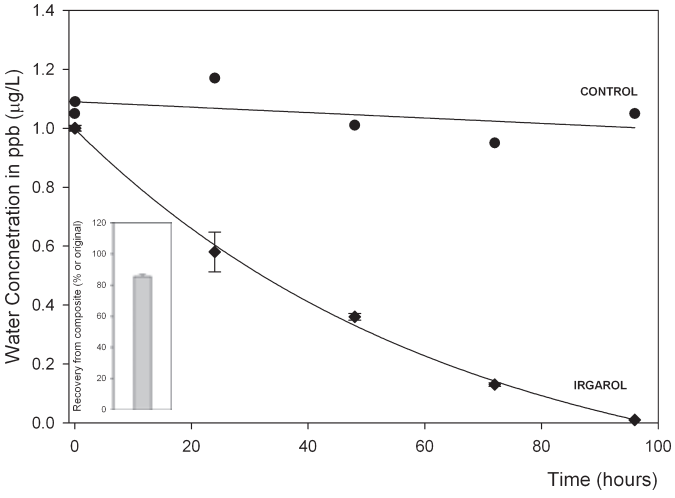


FIG. 8—*Results of the low concentration depletion study of Irgarol 1051 using Fe–PDMS rods showing complete removal after 4 days of exposure and subsequent recovery of the analyte (86 ± 2%) from the Fe–PDMS rods.*

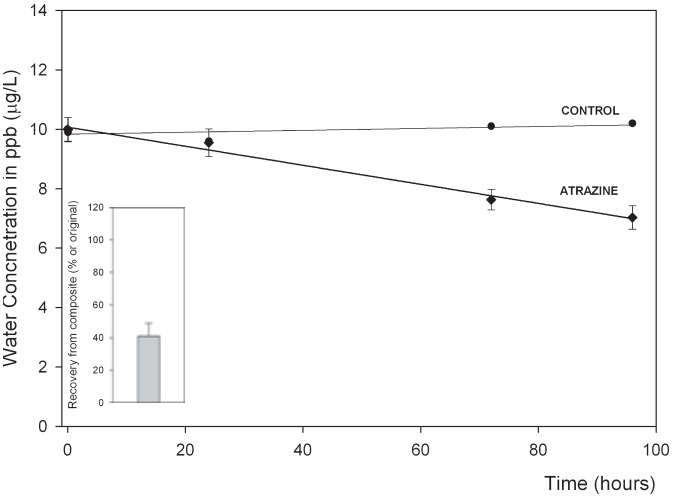


FIG. 10—*Uptake and release of Atrazine from the Fe–PDMS pellets. Initial water concentration was 10 μg/L (●) and remained constant during the exposure experiment. Uptake curve shows removal of only 29.7% of the concentration (◆) after 4 days of exposure. Re-extraction of the target compound from the exposed Fe–PDMS pellets produced only partial recovery (41 ± 8%) of Atrazine based on initial concentrations.*

Fe–PDMS composites to uptake recalcitrant contaminants with large K_{OW} (5–6) from sites contaminated with chlorinated pesticides.

Acknowledgments

The authors of this article would like to acknowledge the instrument and logistic support from the Advanced Mass Spectroscopy Facility and Florida Center for Analytical Electron Microscopy for the use of their scanning electron microscope to obtain images of the Fe–PDMS composites at Florida International University. Funding was received from the Southeast Environmental Research Center (SERC) at Florida International University. This is SERC contribution 412.

References

1. Gang S, Lee HK. Hollow fiber-protected liquid-phase microextraction of triazine herbicides. Anal Chem 2002;74:648–54.
2. Banks KE, Turner PK, Wood SH, Matthews C. Increased toxicity to ceriodaphnia dubia in mixtures of atrazine and diazinon at environmentally realistic concentrations. Ecotoxicol Environ Saf 2005;60:28–36.
3. US Environmental Protection Agency (USEPA). Atrazine reregistration docket, http://www.epa.gov/oppsrrd1/reregistration/atrazine/ (accessed March 4, 2009).
4. Phyu YL, Warne MSJ, Lim RP. Toxicity and bioavailability of atrazine and molinate to the freshwater shrimp (*Paratya australiensis*) under laboratory and simulated field conditions. Ecotoxicol Environ Saf 2005;60:113–22.
5. Gross TS, McCoy KA, Sepulveda M, Carr JA, Giesy JP, Homser AJ, et al. Atrazine exposure and the occurrence of reproductive abnormalities in field caught *Bufo marinus* from south Florida; U.S. Geological Survey Florida Integrated Science Center and University of Florida, http://sofia.usgs.gov/projects/eco_risk/atrazine_geer03abs.html (accessed March 5, 2009).
6. Christopher SV, Bird KT. The effects of herbicides on the development of *Myriophyllum spicatum L.* cultured in vitro. J Environ Qual 1992;21:203–7.
7. Abou-Waly H, Abou-Setta MM, Nigg HN, Mallory LL. Dose-response relationship of *Anabaena flos-aquae* and *Selenastrum capricornutum* to atrazine and hexizinone using chlorophyll-a content and 14C uptake. Aquat Toxicol 1991;20:195–204.
8. Graymore M, Stagbitti F, Allison G. Impacts of atrazine in aquatic ecosystems. Environ Int 2001;26:483–95.
9. Lambropoulou DA, Sakkas VA, Albanis TA. Analysis of antifouling biocides Irgarol 1051 and Sea Nine 211 in environmental water samples using solid-phase microextraction and gas chromatography. J Chromatogr A 2002;952:215–27.
10. Hall LW, Killen WD, Gardinali PR. Occurrence of Irgarol 1051 and its major metabolite in Maryland waters of Chesapeake Bay. Mar Pollut Bull 2004;48:554–62.
11. Gardinali PR, Plasencia MD, Maxey C. Occurrence and transport of Irgarol 1051 and its major metabolite in coastal waters from South Florida. Mar Pollut Bull 2004;49:1072–83.
12. Stuer-Lauridsen F. Review of passive accumulation devices for monitoring organic micropollutants in the aquatic environment. Environ Pollut 2005;136:503–24.
13. Pawliszyn J. Solid phase microextraction: theory and practice. New York, NY: John Wiley & Sons Inc, 1997.
14. Al-Fadul SM. Using magnetic extractants for removal of pollutants from water via magnetic filtration [dissertation]. Stillwater (OK): Oklahoma State University, 2006.

Additional information and reprint requests:
Piero R. Gardinali, Ph.D.
Associate Professor
Department of Chemistry & Biochemistry and
Southeast Environmental Research Center (SERC)
Florida International University, Biscayne Bay Campus
3000 NE 151 St, MSB-356
North Miami, FL 33181
E-mail: Piero.Gardinali@FIU.EDU

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01430.x
Available online at: interscience.wiley.com

PAPER

PATHOLOGY AND BIOLOGY

Morio Iino,[1,2] *M.D., Ph.D. and Chris O'Donnell,*[1] *M.B., B.S., M.Med., Grad.Dip.For.Med.*

Postmortem Computed Tomography Findings of Upper Airway Obstruction by Food

ABSTRACT: This study is a retrospective analysis of 14 cases with food bolus upper airway obstruction as the defined cause of death where both postmortem computed tomography and autopsy were performed. Three groups were defined by the images i.e., Type 1: foreign body situated between the oral cavity and oropharynx, while the epiglottis sits in normal position, Type 2: foreign body situated in the oropharynx just above the epiglottis pushing it posteriorly and obstructing the airway, and Type 3: foreign body obstructing the laryngeal inlet while pushing the epiglottis anteriorly. At the time of autopsy, foreign bodies were detected by pathologists, occasionally in a different position, presumably being dislodged in the act of organ removal especially for the "Type 1" pattern. CT imaging provides accurate interrogation of upper airway bolus obstruction prior to autopsy.

KEYWORDS: forensic science, forensic radiology, computed tomography, postmortem imaging, airway obstruction, food, choking, asphyxia

In choking, asphyxia is caused by obstruction to the air passages. The manner of death can be natural, homicidal, or accidental. Natural deaths are seen in individuals with acute fulminating epiglottitis, laryngeal cyst, or polyp (1–3). Suicidal (4,5) or homicidal (6,7) deaths by choking are relatively uncommon. Most choking deaths are accidental in manner. In children, choking usually involves aspiration of a small object, such as a small rubber ball or a balloon, into the larynx with occlusion of the airway (1,8–11). In adults, choking virtually always involves food. It is commonly associated with acute alcohol intoxication, poor condition or absence of teeth, neurological conditions, or senility (1,12). The piece of food will usually wedge in the laryngopharynx and larynx, completely obstructing the airway. It is commonly known as the "café coronary" whereby sudden and unexpected death occurs during a meal caused by accidental occlusion of the airway by food (1,12,13).

The diagnosis of choking death by food is made at autopsy when the airway is found to be occluded. The finding of food material in the airway at autopsy alone does not indicate that the individual has choked to death. Approximately 20–25% of all individuals have agonal aspiration of gastric contents, independent of the cause of death. If the individual had an occluded airway and the object or food was removed during resuscitation, the only way to make the diagnosis would be by review of the clinical history (1).

Computed tomography (CT) in the clinical setting of upper esophageal obstruction is a suitable way for detection of foreign bodies (14,15). There are now a number of centers in the world investigating the application of CT and magnetic resonance imaging (MRI) in the forensic environment, and most forensic institutes will seek regular access to such CT facilities or install machines into their own mortuaries in the next 5–10 years (16). Indeed, at one of those centers both MRI and CT have recently been shown to accurately detect laryngeal foreign bodies in three postmortem cases (17). In this study, we have evaluated the use of postmortem CT imaging in a larger cohort of deceased persons with upper airway obstruction by food stuff.

Materials and Methods

A retrospective study of autopsy files at Victorian Institute of Forensic Medicine (VIFM) was undertaken for all cases of death caused by upper airway obstruction by food over a 3½-year period from April 2005 to October 2008. VIFM plays an important role in death investigation being an integral part of the coronial system of Victoria, an Australian state with a population of *c.* 5.2 million people. Approximately 3000 coronial autopsies are performed at the Institute each year.

A total of 26 deaths were determined at autopsy to be attributed to laryngeal obstruction by food. Cases excluded from the sample were those in which food in the airway had been removed at the scene ($n = 6$), cases that died in hospital with brain hypoxia or multiple organ failure days after the incidents ($n = 5$), and one where the CT scan was not performed for unknown reasons ($n = 1$). Individuals with aspiration of food or vomitus beyond the vocal chords into the tracheobronchial tree alone were not included in the case search.

All deceased persons admitted to the Institute since April 2005 have been CT scanned from the top of the head to toes prior to autopsy or external examination. A 16-channel multi-detector CT scanner (Aquilion16; Toshiba Medical Systems Corporation, Tochigi, Japan) was used to take multiplanar reconstruction (MPR)

[1]Victorian Institute of Forensic Medicine, Department of Forensic Medicine, Monash University, Southbank, Vic. 3006, Australia.
[2]Department of Legal Medicine, Osaka University Graduate School of Medicine, Osaka 560-0871, Japan.

Received 20 May 2009; and in revised form 29 June 2009; accepted 4 July 2009.

images including axial, coronial, and sagittal sections. It also provides 3D images such as volume rendering and maximum intensity projection.

The head and body were scanned separately, and images reconstructed into overlapping 1-mm and 2-mm slices, respectively. Images were sent to the Institute's picture archiving and communication systems server and analyzed on a workstation (Vitrea 2; Vital Images, Minnetonka, MN).

Each case was subjected to a complete autopsy. Toxicology was conducted on most cases, except for numbers 3, 4, 9, and 13. Toxicological testing included analysis for alcohol, drugs of abuse (amphetamines and related stimulants, benzodiazepines, cannabinoids, cocaine, and opiates), and a large range of common prescription and over-the-counter drugs. Ethanol was quantified by gas chromatography. All other detected drugs were confirmed and quantified using appropriate mass spectrometric techniques.

This study was approved by the ethics committee of VIFM.

Results

There were 14 cases (seven men) of upper airway obstruction by food detected at autopsy. Median age was 64.5 ranging from 46 to 84. Details of the cases are shown in Table 1.

Incidents leading to death occurred at home ($n = 7$) or care facilities ($n = 7$) including nursing home, psychiatric facility, or retirement home. In six of seven cases in which the incidents took place at home, the deceased were found on the kitchen floor. In four of seven cases that took place in care facilities, the incidents happened during meal time. No cases took place outside a residential area such as a café or restaurant.

At least eight of the cases had known neurological conditions or psychiatric problems such as schizophrenia, dementia, or encephalopathy.

Alcohol and/or drugs were detected in nine of the cases. In five cases (Cases 2, 5, 6, 7, and 10), alcohol was detected at high concentrations (blood alcohol concentration—BAC > 0.1 g/100 mL), and all of these had a history of alcoholism. The results of other toxicological findings are shown in Table 1.

Food items obstructing the airways were detected at autopsy in all cases. The materials found were listed by the pathologists as unspecified ($n = 5$), bread ($n = 3$), meat ($n = 3$), cake ($n = 1$), "dim sim" (meat dumpling, $n = 1$), and grape ($n = 1$).

Radiologically all food in the upper airways was detected on CT.

Case 1

Case History

A 57-year-old woman was found on the kitchen floor of her home during the daytime. The deceased had a history of bowel cancer that had been treated surgically and also breast cancer. She suffered from a spasmodic twitch and depression, having received electro-convulsive therapy in the past.

Radiological Findings

CT scanning of the deceased was performed *c*. 15 h after death was certified. An axial image of the upper airway at the level of the mandible shows airway obstruction by a rounded foreign body measuring 2 × 2 cm (Fig. 1*a*).

Sagittal MPR confirms the rounded foreign body in the naso-oro-pharynx (Fig. 1*b*).

Autopsy Findings

A partly chewed but still essentially ovoid grape was situated at and obstructing the laryngeal aperture. It weighed 14 g and measured 4 × 3 × 2 cm.

Toxicology Findings

Citalopram and olanzapine in the blood were analyzed by capillary gas chromatography with nitrogen phosphorous and mass spectrometry (GC-B/MS) and detected at levels of *c*. 0.8 and 0.5 mg/L, respectively.

Case 2

Case History

A 62-year-old man was found on the kitchen floor of his home during daytime. The deceased was a heavy drinker and had, on multiple occasions, choked on food while in a state of intoxication.

Radiological Findings

CT scan of the deceased was performed *c*. 14 h after the estimated time of death. Axial image of the upper airway shows partial obstruction by foreign body (Fig. 2*a*). The sagittal section demonstrates an oval-shaped foreign body measuring 1.5 × 1 cm in the oropharynx immediately above the epiglottis displacing it posteriorly (Fig. 2*b*). It also shows further solid material in the naso-oropharynx (Fig. 2*b*).

Autopsy Findings

An amorphous fragment of coarse pale food measuring *c*. 2.5 × 2 cm was located immediately superior to the true vocal cords. A moderate amount of coarse and finely divided food particles was also identified within the distal trachea and within the right main bronchus.

Toxicology Findings

Toxicological analysis of body fluids disclosed alcohol in blood and vitreous at concentrations of 0.25 and 0.33 g/100 mL, respectively.

Case 3

Case History

An 84-year-old female resident of an aged care home with history of dementia was given a piece of sponge cake by another resident's family. She appeared to choke on the food, gasping for breath, and waving her hand in front of her face. Staff attempted a Heimlich maneuver and used a suction device in an attempt to dislodge the food item, but these proved unsuccessful. She had a history of choking on four occasions in the past 9 months.

Radiological Findings

CT scan of the deceased was performed *c*. 4 h after death was certified. Axial images show upper airway obstruction by a rounded foreign body measuring 2 × 2 cm (Fig. 3*a*). The foreign body has a high X-ray attenuation measuring up to 220 Hounsfield Units

TABLE 1—*Cases of upper airway obstruction by food.*

Case	Age	Sex	Place of the Incident	Medical Conditions	Food Items	Positions of Foreign Materials Seen on CT Images	Positions of Foreign Materials Found at Autopsy	Results of Toxicology		Number of Teeth		Classification of the Choking Types by CT Images‡
								Blood concentration	Therapeutic level	Upper	Lower	
1	57	F	Kitchen (home)	Breast cancer Spasmodic twitch	Grape	Naso-oro-pharynx	Larynx	Citalopram: 0.8 mg/L Olanzapine: ~0.5 mg/L	Up to ~0.6mg/L Up to ~0.5mg/L	0	10	Type 1
2	62	M	Kitchen (home)	Alcoholism	Unspecified	Naso-oro-pharynx and oropharynx	Oropharynx Distal trachea Right main bronchus	Blood alcohol: 0.25 g/100 mL Diazepam: ~0.5 mg/L	0.7–1.5 mg/L	14	10	Type 1 + 2
3	84	F	Nursing home	Dementia	Cake	Laryngopharynx	Supralaryngeal region	–†		0	0	Type 3
4	67	M	Kitchen (home)	Inner ear disorder	Meat	Naso-oro-pharynx and laryngopharynx	Upper oesophagus and the larynx	–†		4	7	Type 1 + 3
5	69	M	Kitchen (home)	Alcoholism Stroke	Meat	From the naso-oro-pharynx to the laryngopharynx	Posterior oropharynx	Blood alcohol: 0.27 g/100mL		2	2	Type 3
6	46	M	Kitchen (home)	Alcoholism Hepatitis C	Steamed Dim-Sim*	From the oropharynx to the laryngopharynx	Larynx	Blood alcohol: 0.29 g/100 mL		0	5	Type 3
7	57	F	Bathroom (home)	Alcoholism Psychiatric disease	Chicken	From the oropharynx to the laryngopharynx	Larynx	Blood alcohol: 0.14 g/100 mL Diazepam: 0.2 mg/L Olanzapine: 0.4 mg/L Venlafaxine: ~0.2 mg/L	0.7–1.5 mg/L Up to ~0.5 mg/L Up to ~0.5 mg/L	13	12	Type 3
8	76	F	Care facility	Wernike's encephalopathy secondary to ethanol abuse	Bread	From the naso-oro-pharynx to the laryngopharynx and tracheal bifurcation	Tracheal lumen	Imipramine: 0.8 mg/L Desipramine: ~1.1 mg/L	Plasma concentrations of imipramine and desipramine vary widely between individuals (20).	0	8	Type 3
9	83	M	Nursing home	Chronic obstructive airways disease Pneumonia	Unspecified	Entire oesophagus, the laryngopharynx, trachea and bronchi	Entire oesophagus, the laryngopharynx, trachea and bronchi	–†		8	10	Type 3
10	61	M	Kitchen (home)	Alcoholism	Unspecified	From the oropharynx to the laryngopharynx	Larynx	Blood alcohol: 0.2 g/100 mL		4	5	Type 1 + 3
11	48	F	Care facility	Schizophrenia	Bread	From the oropharynx to the laryngopharynx	Larynx	Olanzapine: ~0.3 mg/L	Up to ~0.5 mg/L	11	13	Type 3
12	80	F	Retirement home	Unknown	Unspecified	Laryngopharynx	Larynx	Negative		0	6	Type 3
13	81	M	Nursing home	Schizophrenia	Unspecified	From the oropharynx to the laryngopharynx	Upper esophagus and laryngopharynx	–†		0	0	Type 3
14	59	F	Psychiatric facility	Schizophrenia	Bread	From the oropharynx to the laryngopharynx	Larynx	Valproic acid: 10 mg/L Chlorpromazine: 0.1 mg/L	50 to 100 mg/L Up to ~1 mg/L	0	11	Type 3

*Chinese-inspired meat dumpling-style snack food.
†No toxicological analysis has been carried out.
‡Details of classification are shown in Table 2.

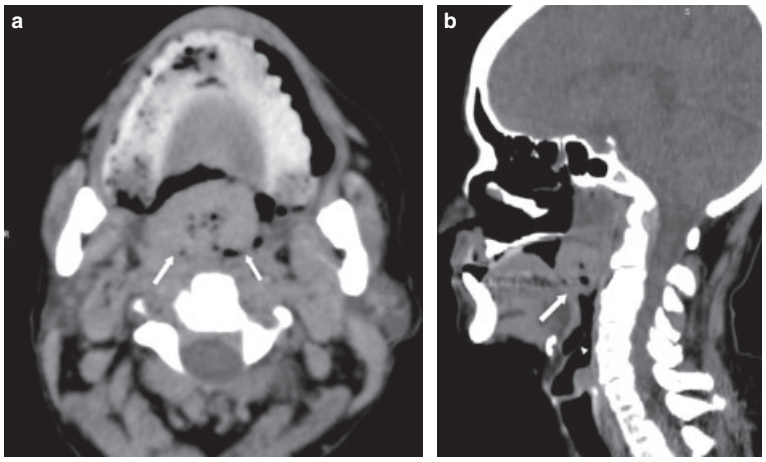


FIG. 1—*(a) Axial image of the upper airway shows airway obstruction by food (a grape, arrows). (b) Sagittal image shows the round shaped grape in the naso-oro-pharynx (arrow). The position of the epiglottis is not affected by food (arrow head).*

(HU). The sagittal section and 3D volume-rendered image clearly demonstrate the ball-shaped foreign body sitting behind the epiglottis in the laryngopharynx inlet (Fig. 3*b,c*). The volume-rendered 3D image shows the absence of teeth (Fig. 3*c*).

Autopsy Findings

There was soft somewhat mushy, pink food material situated behind the epiglottis causing obstruction in the supralaryngeal region (Fig. 3*d*).

Case 4

Case History

A 67-year-old man was found in the doorway of his kitchen/dining room at home. It appeared that he was attempting to make a sandwich in the kitchen. He was in good health apart from an inner ear disorder for which he was taking medication.

Radiological Findings

CT scan of the deceased was performed *c.* 84 h after death was certified. Axial images show upper airway obstruction by foreign body (Fig. 4*a*).

The sagittal section demonstrates two pieces of foreign body, one in the naso-oro-pharynx and one in the laryngopharynx just behind the epiglottis (Fig. 4*b*). Coronal image shows the shapes of these foreign bodies (Fig. 4*c*).

Autopsy Findings

Two pieces of meat were found obstructing the larynx. The first was situated in the upper esophagus and epiglottic region measuring $3.5 \times 2 \times 1.5$ cm, while the second longer piece extended down into the larynx. This measured $6 \times 2 \times 0.7$ cm (Fig. 4*d*).

Extra Case

Case History

A 37-year-old woman was found in bed by her father. The deceased had a long history of schizophrenia and sleep apnea.

Radiological Findings

CT scan of the deceased was performed *c.* 12 h after death was certified. The sagittal section shows oval-shaped material in the oropharynx just above the epiglottis (Fig. 5).

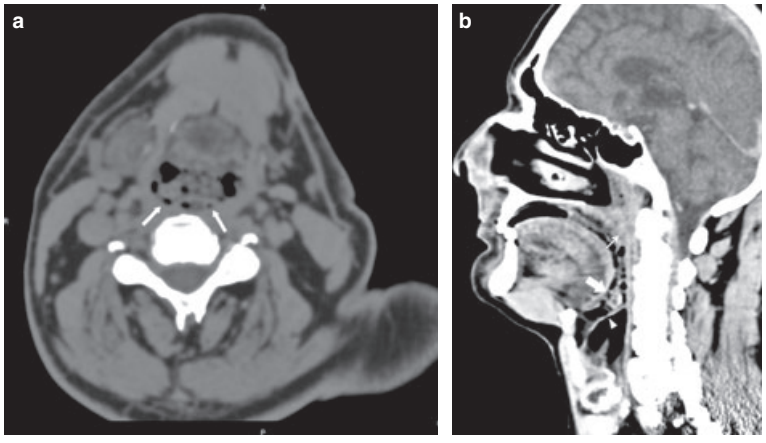


FIG. 2—*(a) Axial image of the upper airway shows airway obstruction by unspecified food (arrows). (b) Sagittal image shows the oval shaped foreign body in the oropharynx immediately above the epiglottis (large arrow). Note there is also further foreign material in the naso-oro-pharynx (small arrow). The epiglottis is pushed down (arrow head).*

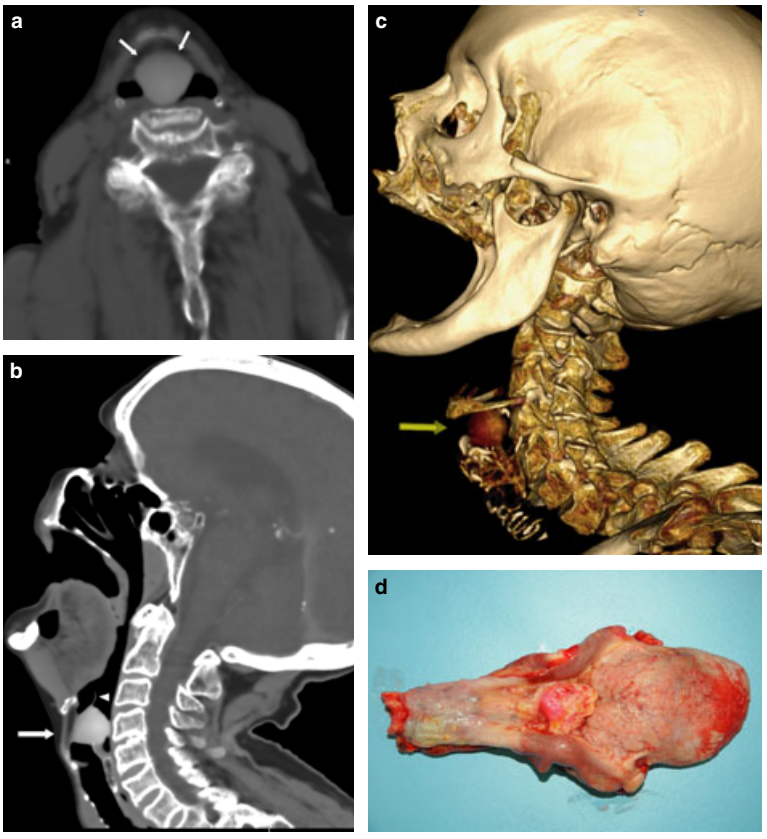


FIG. 3—*(a) Axial image of the upper airway shows airway obstruction by high density round shaped food (cake, arrows). (b) Sagittal image shows the ball shaped cake in the laryngopharynx behind the epiglottis (arrow). The epiglottis is pushed forward (arrow head). (c) Three-dimensional volume rendered image demonstrates the cake in the laryngopharynx (arrow). (d) A pink cake piece found at autopsy.*

Autopsy Findings

Prominent lymphoid hyperplasia (enlarged tonsils) was detected in the oropharynx.

No foreign material was present within the pharynx or tracheal lumen. The coronary arteries show minimal atheroma; however, the left circumflex coronary artery, *c.* 1 cm from the vessel origin, showed a fibrofatty atheromatous plaque with up to 75% luminal stenosis.

Discussion

There are several risk factors for choking on food. These include middle-aged men who eat large food boluses, with poor condition or absence of the teeth, and alcohol intoxication (12). Individuals with neurological conditions such as atherosclerotic cerebrovascular disease, dementia, Parkinson's disease, and psychiatric disorders such as schizophrenia are also at risk (12,18).

In our cases, five of seven men (71%) were middle aged in the range of 40–70 years (Cases 2, 4, 5, 6, and 10). Four of them (80%) had poor condition of teeth i.e., ≤4 teeth in the upper jaw (Cases 4, 5, 6, and 10). Four (80%) also had a history of alcoholism with significantly elevated blood alcohol levels (BAC > 0.20 g/100 mL) at the time of death (Cases 2, 5, 6, and 10). This would have depressed the central nervous system with reduced reflexes thus contributing to the lodgment of food in the airways (19).

There were at least eight cases (57%) that had known history of neurological conditions (n = 4) or psychiatric disorders (n = 4). In five other cases, prescribed drugs were detected in concentrations consistent with therapeutic use (Cases 1, 2, 7, 11, and 14). These included an anticonvulsant (Case 1), antidepressants (Case 7), and antipsychotics (Cases 7 and 14). In one case, therapeutic ranges for imipramine and desipramine have not been defined because the plasma concentrations vary widely between individuals (Case 8, Table 1) (20).

All foreign bodies were identified on MPR CT images even in a severely decomposed body (Case 10). Most were difficult to detect on 3D volume-rendered images, although teeth were readily identified in the jaw. The CT images sometimes provided an estimation of the type of food material present. For example, images of unchewed meat in Case 4 showed a striped pattern suggestive of muscle (Fig. 4*a,b,c*), while a piece of cake in Case 3 was shown to be hyperdense (Fig. 3*a,b,c*).

In this study, we noticed three different CT patterns of laryngeal obstruction (Table 2). The first (Type 1) is where the foreign body sits between the oral cavity and oropharynx with the epiglottis in its normal position (Figs. 1*b* and 6*a*). In this type, the foreign material obstructs by sitting at the junction of oral and nasal airways. The foreign body does not reach or make contact with the epiglottis. The second (Type 2) is where the foreign body is situated in the oropharynx just above the epiglottis. In this position, the epiglottis is pushed posteriorly by the foreign body leading to airways obstruction (Figs. 2*b* and 6*b*). The last one (Type 3) is where the foreign body is situated in the laryngeal inlet just above the vocal cord. In this position, it obstructs the laryngeal airway while displacing the epiglottis forward. If large enough, the foreign body may also extend into the oropharynx and even the nasopharynx

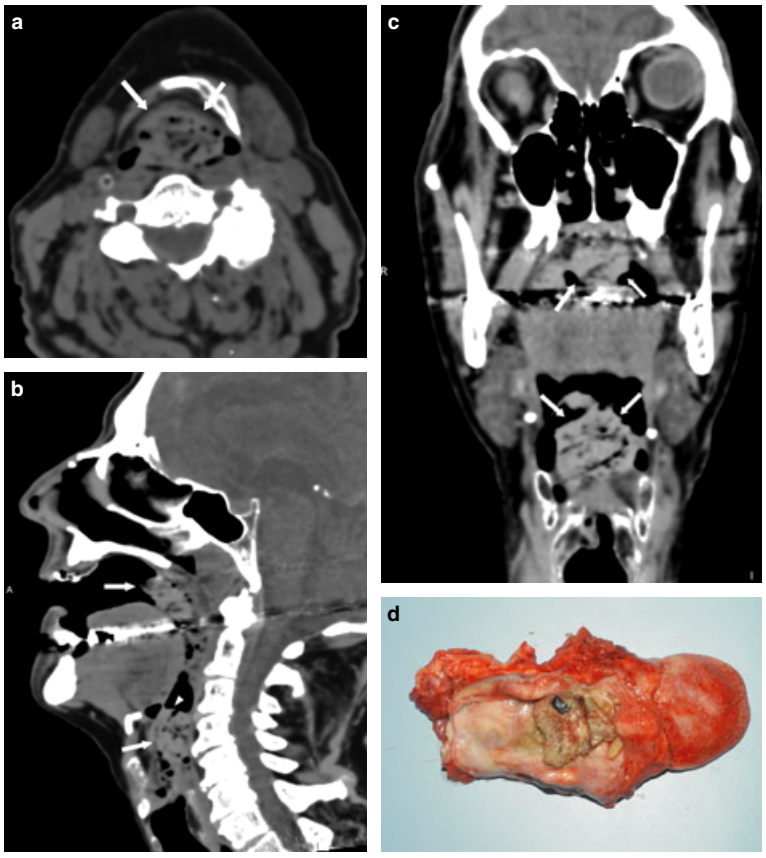


FIG. 4—*(a) Axial image shows upper airway obstruction by foreign body (meat, arrows). (b) The sagittal section demonstrates two pieces of meat, one in the naso-oro-pharynx (upper arrow) and one in the laryngopharynx just behind the epiglottis (lower arrow). The epiglottis is pushed forward (arrow head). (c) Coronal image shows the shapes of two pieces of meat (arrows). (d) A piece of meat found in the laryngopharynx at autopsy.*

FIG. 5—*The sagittal section shows oval-shaped "material" (enlarged tonsil, arrow) in the oropharynx above the epiglottis (arrow head).*

(Figs. 3*b* and 6*c*). In the 14 cases, we had 10 cases of Type 3 and one case of Type 1. In the other three cases, two were combined Types 1 and 3, while one was a combined Types 1 and 2 (Table 1).

Location of foreign bodies need not necessarily be the same at autopsy as at the time of death. Foreign material can be dislodged during organ removal. Material in the oropharynx for example may move on extraction of the tongue. Pathologists therefore need to be careful when diagnosing airway obstruction by food in the laryngo-pharynx as it may have "dropped down" from the oropharynx. CT is useful therefore as it provides an overview of the airway prior to any intervention. It can be viewed preceding the autopsy for procedure planning or reviewed subsequent to autopsy if the pathologist needs confirmation of foreign body location.

In Case 1, the pathologist at autopsy diagnosed food, a grape, obstructing the laryngeal aperture, although it was trapped in the naso-oro-pharynx when the body was scanned on CT (Fig. 1*b*). In Case 8, the pathologist detected foreign material only in the tracheal lumen at autopsy, but at the time of the CT scan it was situated from the naso-oro-pharynx to the laryngopharynx (Table 1).

The key to radiological evaluation of the upper airway is interrogation of the sagittal MPR view. It must be reconstructed in a midline plane orthogonal to the long axis of the pharyngo-larynx. This is not always easy especially in cases where the head and body are distorted because of rigor. It may require sophisticated workstation manipulation. The epiglottis may also be difficult to detect in the sagittal view. Any finding that is made on the sagittal view must be confirmed by referring to the corresponding axial image.

TABLE 2—*Classification of choking based on the postmortem CT findings of upper airway.*

	Type 1	Type 2	Type 3
Position of food	In the naso-oro-pharynx	In the oropharynx above the epiglottis	In the laryngopharynx behind the epiglottis
Position of epiglottis	Normal	Pushed down by food	Pushed forward by food
Figures	Fig. 6*a*	Fig. 6*b*	Fig. 6*c*

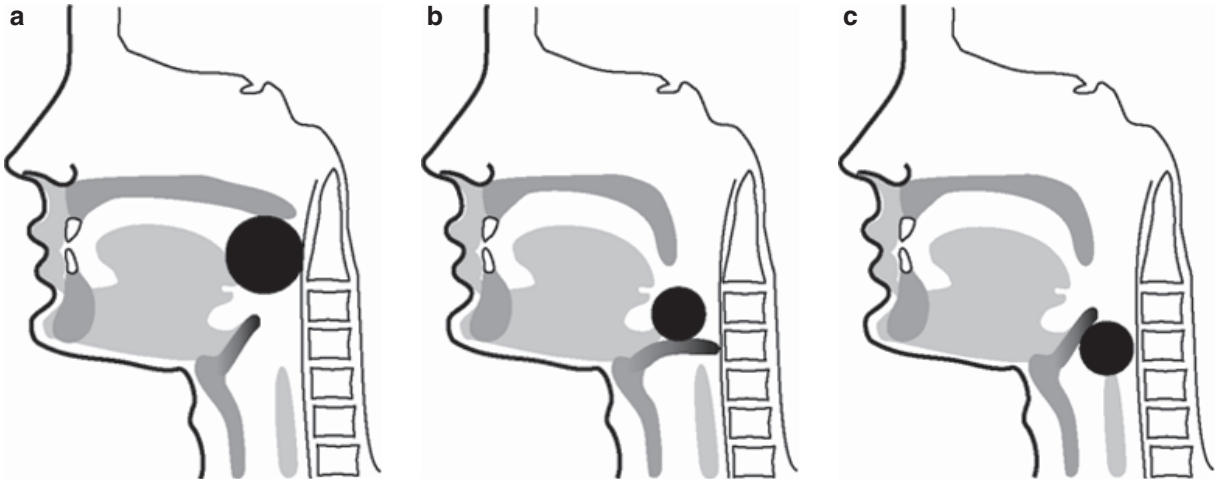


FIG. 6—*Classification of choking based on the postmortem CT findings of upper airway obstruction. (a) Type 1: The food is situated in the naso-oro-pharynx while the position of the epiglottis has not been affected. (b) Type 2: The food is situated above the epiglottis causing the epiglottis to be pushed down by the food. (c) Type 3: The food is situated in the laryngeal inlet by pushing the epiglottis forward.*

A possible false positive for upper airway obstruction by food is airway obstruction because of organ enlargement such as the tonsils or even malignancy. In our Extra Case, the sagittal CT section (Fig. 5) produced an image very similar to the image of "Type 1" (Fig. 1*b*) with "material" in the oropharynx above the epiglottis. Autopsy revealed enlarged tonsils. It is very important therefore to distinguish between anatomical/pathological enlargement and foreign material on postmortem CT images and if necessary view the airway directly for confirmation. In one reported case, postmortem CT angiography has been shown to be helpful in discriminating between foreign body and soft tissue mass in the larynx by the demonstration of blood vessels in the mass (17) although this technique is not yet widely available.

Another potential false positive is food or gastric content that has been regurgitated from the esophagus and stomach into the upper airway around the time of death. This can produce a CT appearance of upper airway foreign material that may be indistinguishable from our findings in laryngeal obstruction. Correlation with the circumstances surrounding death is therefore important in making a definitive diagnosis of choking.

At VIFM, CT has been shown to be a very useful tool to assist forensic pathologists in cases of laryngeal obstruction by food. Section 29 of the Coroners Act 1985 (21) in Victoria, Australia allows families to object to autopsy. CT in the appropriate clinical setting can confirm the diagnosis of laryngeal obstruction by food bolus or where there is no obvious cause of death provide some indication of that diagnosis to the pathologist.

Acknowledgments

We are grateful to the pathologists and toxicologists at VIFM for providing us with information for the cases in detail and also thankful to Sarn Potter for drawing the beautiful illustrations.

References

1. Di Maio VJ, Di Maio D. Asphyxia. In: Di Maio DJ, Di Maio D, editors. Forensic pathology, 2nd edn. New York, NY: CRC Press LLC, 2001;229–77.
2. Porzionato A, Macchi V, Rodriguez D, De Caro R. Airway obstruction by laryngeal masses in sudden and homicidal. Forensic Sci Int 2007; 171: e15–20.
3. Carrick C, Collins KA, Lee CJ, Prahlow JA, Barnard JJ. Sudden death due to asphyxia by esophageal polyp: two case reports and review of asphyxial deaths. Am J Forensic Med Pathol 2005;26:275–81.
4. Sauvageau A, Yesovitch R. Choking on toilet paper: an unusual case of suicide and a review of the literature on suicide by smothering, strangulation, and choking. Am J Forensic Med Pathol 2006;27:173–4.
5. Pampin JB, Varela LG. Suicidal choking caused by a bizarre combination of inhalation to the bronchi and external neck compression. Leg Med 2001;3:119–22.
6. Cohle SD. Fatal pepper aspiration. Homicidal asphyxia by pepper aspiration. J Forensic Sci 1986;31:1475–8.
7. Kurihara K, Kuroda N, Murai T, Shinozuka T, Yanagida J, Matsuo Y, et al. A case of homicidal choking mistaken for suicide. Med Sci Law 1992;32:65–7.
8. Zarrin-Khameh N, Lyon RE. Images in clinical medicine. Asphyxia an inhaled foreign body. N Engl J Med 2005;352:2110.
9. Abdel-Rahman HA. Fatal suffocation by rubber balloons in children: mechanism and prevention. Forensic Sci Int 2000;108:97–105.
10. Jumbelic MI. Airway obstruction by a ball. J Forensic Sci 1999;44: 1079–81.
11. Meel BL. An accidental suffocation by a rubber balloon. Med Sci Law 1998;38:81–2.
12. Wick R, Gilbert JD, Byard RW. Café coronary syndrome—fatal choking on food: an autopsy approach. J Clin Forensic Med 2006;13:135–8.
13. Haugen RK. The café coronary. Sudden death in restaurants. JAMA 1963;186:142–3.

14. Braverman I, Gomori JM, Polv O, Saah D. The role of CT imaging in the evaluation of cervical esophageal foreign bodies. J Otolaryngol 1993;22:311–4.
15. Kobayashi M, Seto A, Nomura T, Yoshida T, Yamamoto M. 3D-CT highly useful in diagnosing foreign bodies in the paraesophageal orifice. Nippon Jibiinkoka Gakkai Kaiho 2004;107:800–3.
16. O'Donnell C, Woodford N. Post-mortem radiology—a new sub-specialty? Clin Radiol 2008;63:1189–94.
17. Oesterhelweg L, Bollinger S, Thali M, Ross S. Virtopsy: postmortem imaging of laryngeal foreign bodies. Arch Pathol Lab Med 2009;133:806–9.
18. Ruschena D, Mullen PE, Palmer S, Burgess P, Cordner SM, Drummer OH, et al. Choking deaths: the role of antipsychotic medication. Br J Psychiatry 2003;183:446–50.
19. Di Maio VJ, Di Maio D. Interpretive toxicology: drug abuse and drug deaths. In: Di Maio DJ, Di Maio D, editors. Forensic pathology, 2nd edn. New York, NY: CRC Press LLC, 2001;507–45.
20. Imipramine. Martindale: the complete drug reference, 35th edn. London: Pharmaceutical Press, 2007.
21. Coroners Act 1985. S.29 Objections to autopsy, 1985.

Additional information and reprint requests:
Morio Iino, M.D., Ph.D.
Department of Legal Medicine
Osaka University Graduate School of Medicine
2-2-F3 Yamadaoka, Suita
Osaka 565-0871
Japan
E-mail: iino@legal.med.osaka-u.ac.jp

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01459.x
Available online at: interscience.wiley.com

PAPER

PATHOLOGY/BIOLOGY

Anny Sauvageau,[1,2] *M.D., M.Sc. and Elie Boghossian,*[1] *B.Sc.*

Classification of Asphyxia: The Need for Standardization

ABSTRACT: The classification of asphyxia and the definitions of subtypes are far from being uniform, varying widely from one textbook to another and from one paper to the next. Unfortunately, similar research designs can lead to totally different results depending on the definitions used. Closely comparable cases are called differently by equally competent forensic pathologists. This study highlights the discrepancies between authors and tries to draw mainstream definitions, to propose a unified system of classification. It is proposed to classify asphyxia in forensic context in four main categories: suffocation, strangulation, mechanical asphyxia, and drowning. Suffocation subdivides in smothering, choking, and confined spaces/entrapment/vitiated atmosphere. Strangulation includes three separate forms: ligature strangulation, hanging, and manual strangulation. As for mechanical asphyxia, it encompasses positional asphyxia as well as traumatic asphyxia. The rationales behind this proposed unified model are discussed.

KEYWORDS: forensic sciences, asphyxia, suffocation, strangulation, chemical asphyxia, smothering, choking, mechanical asphyxia, positional asphyxia, traumatic asphyxia

Etymologically, the term asphyxia is derived from the Greek and means "a stopping of the pulse" (1,2). Nowadays, this broad term is used to encompass all conditions caused by the failure of cells to receive or utilize oxygen (1–4). This deprivation can be partial (hypoxia) or total (anoxia) (2). The brain is particularly sensitive to the lack of oxygen and is the most affected organ (3).

Asphyxial deaths are common in forensic practice. In a 21-year study by Azmak, these deaths represented 15.7% of forensic autopsies (5). It should be mentioned however that in the traditional forensic pathology context, the term asphyxia is often used by convention to include conditions that may not be truly asphyxial in nature (4,6).

Unfortunately, the classification of asphyxia and the definition of subtypes are far from being uniform, varying widely from one textbook to another and from one paper to the next. This study will first highlight the discrepancies between authors and then try to draw mainstream definitions, to propose a unified system of classification.

Definition of Asphyxia in the Forensic Context

Asphyxia is often defined as failure of cells to receive or utilize oxygen. It may be argue that this definition is not very useful, at least for a forensic pathology nosologic classification. Alternative definitions may be considered: noncardiac deaths, lack of oxygen delivery or utilization, lack of oxygen delivery to the lung alveoli, lack of oxygen delivery to the lung with an obstruction above the carina. None of these definitions is perfect in the modern forensic practice. Some definitions are too large, encompassing medical conditions such as status asthmaticus, tension pneumothorax, or sickle cell crisis that are not generally considered as belonging to the classification of forensic asphyxia. Others seem too specific and it could be argue that they exclude conditions such as strangulation or drowning. To avoid these two opposite drawbacks, it is recommended to limit the scope of the definition to the forensic settings by simply adding this limitation to the definition itself, while keeping the rest of definition large enough to encompass all entities traditionally include in this category by major textbooks: asphyxia in the forensic context is defined as forensic situations where a body does not receive or utilize adequate amounts of oxygen.

Classification of Asphyxia: An Overview

In forensic textbooks, widely different classifications are found (Fig. 1). In the textbook by DiMaio and DiMaio (2), asphyxial deaths are divided into three broad groups: suffocation, strangulation, and chemical asphyxia (Fig. 1*A*). A very similar classification is presented in a forensic neuropathology textbook by Oehmichen et al. (7) the only difference being the addition of drowning as a form of suffocation (Fig. 1*B*). In an article by Azmak, the DiMaio and DiMaio's classification was also used, but with the addition of drowning as a fourth group of asphyxia instead of as a subtype of suffocation (Fig. 1*C*) (5). In the textbook by Shkrum and Ramsay (1), a classification based on the level of obstruction in mechanical asphyxia is proposed (Fig. 1*D*). The latter differs from the DiMaio and DiMaio's inspired classifications. The most obvious disparity is in the concept of mechanical asphyxia: this term is used in a very restrictive sense by DiMaio and DiMaio, referring to a form of asphyxia by pressure on the outside of the body preventing respiration, whereas the same term is used in a larger way by Shkrum and Ramsay, encompassing all forms of asphyxia in which interference with oxygen and carbon dioxide exchange is caused by

[1]Laboratoire de sciences judiciaires et de médecine légale, 1701 Parthenais street, Montreal, QC, Canada H2K 3S7.
[2]Office of the Chief Medical Examiner, 7007 116 street, Edmonton, AB, Canada T6H 5R8.

Received 6 May 2009; and in revised form 16 July 2009; accepted 7 Sept. 2009.

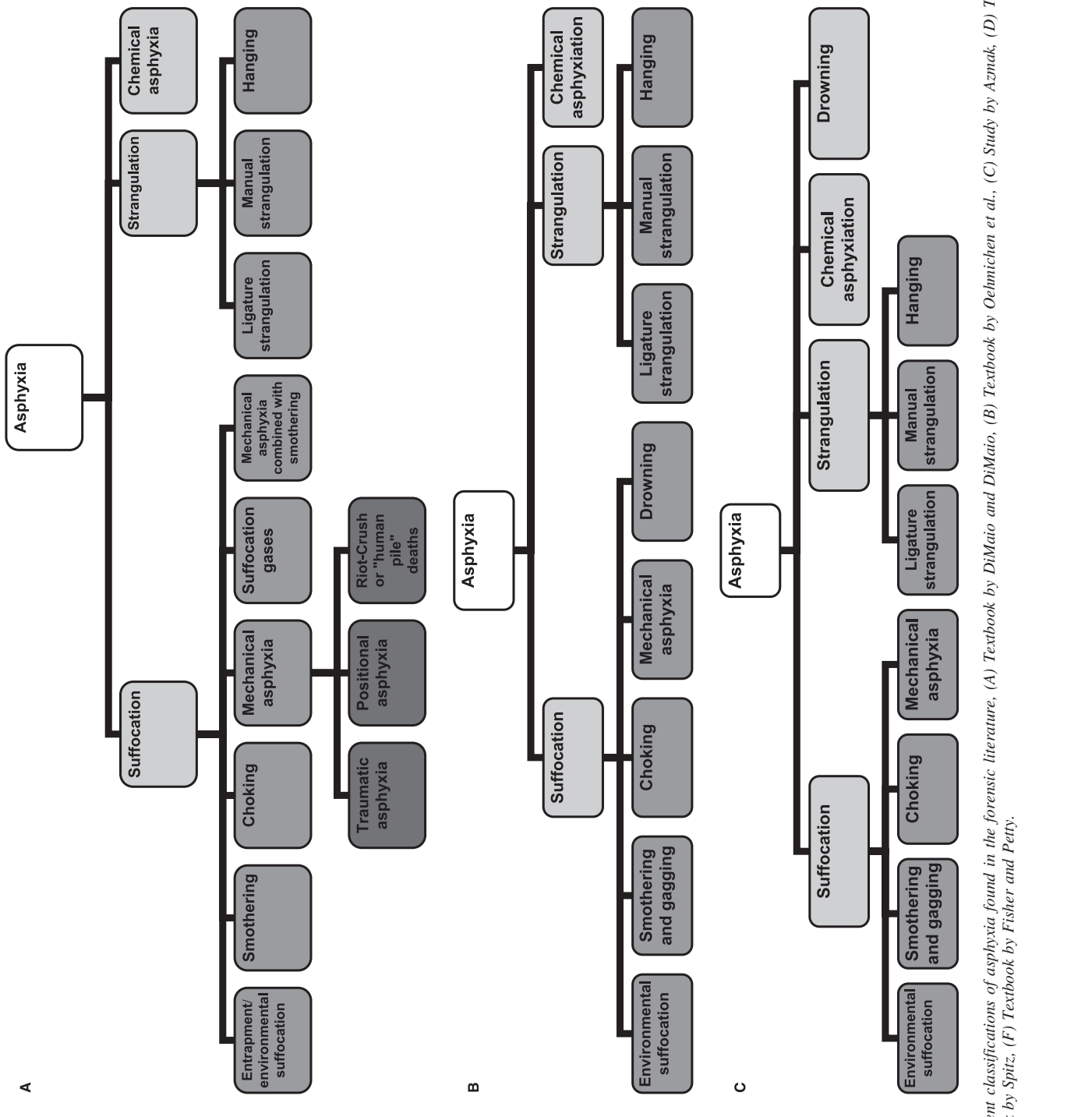
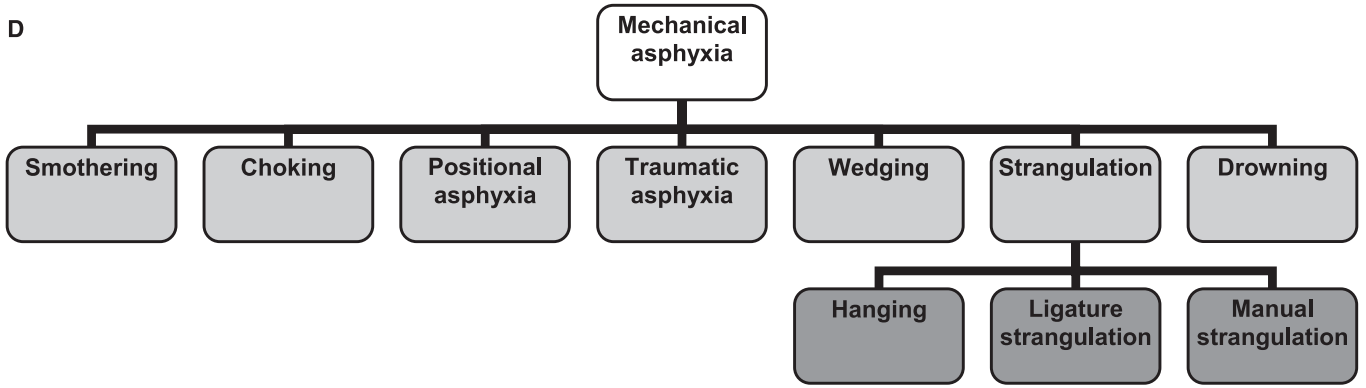
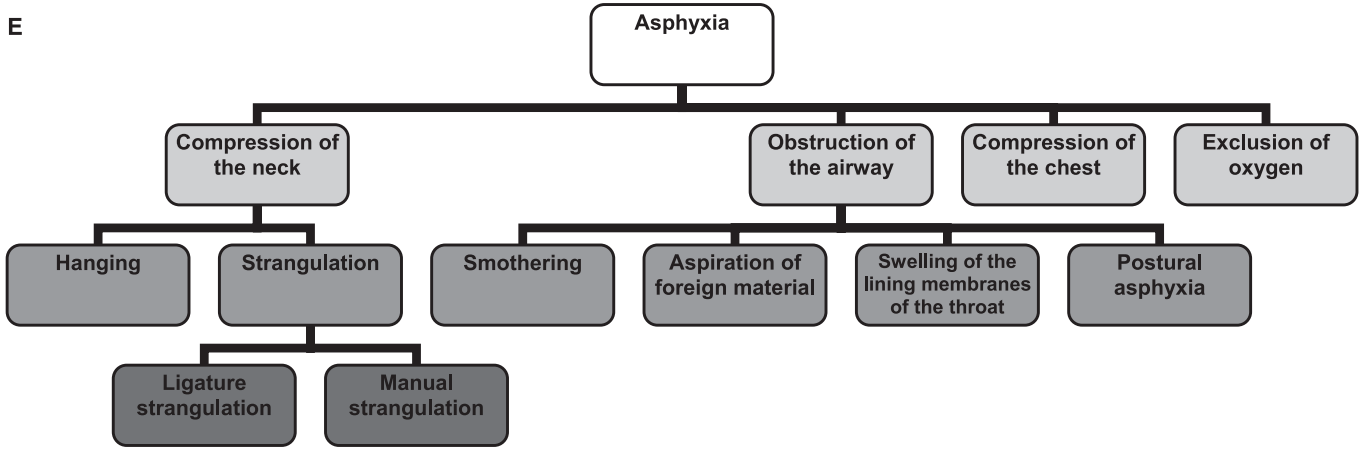


FIG. 1—*The different classifications of asphyxia found in the forensic literature, (A) Textbook by DiMaio and DiMaio, (B) Textbook by Oehmichen et al., (C) Study by Azmak, (D) Textbook by Shkrum and Ramsay, (E) Textbook by Spitz, (F) Textbook by Fisher and Petty.*





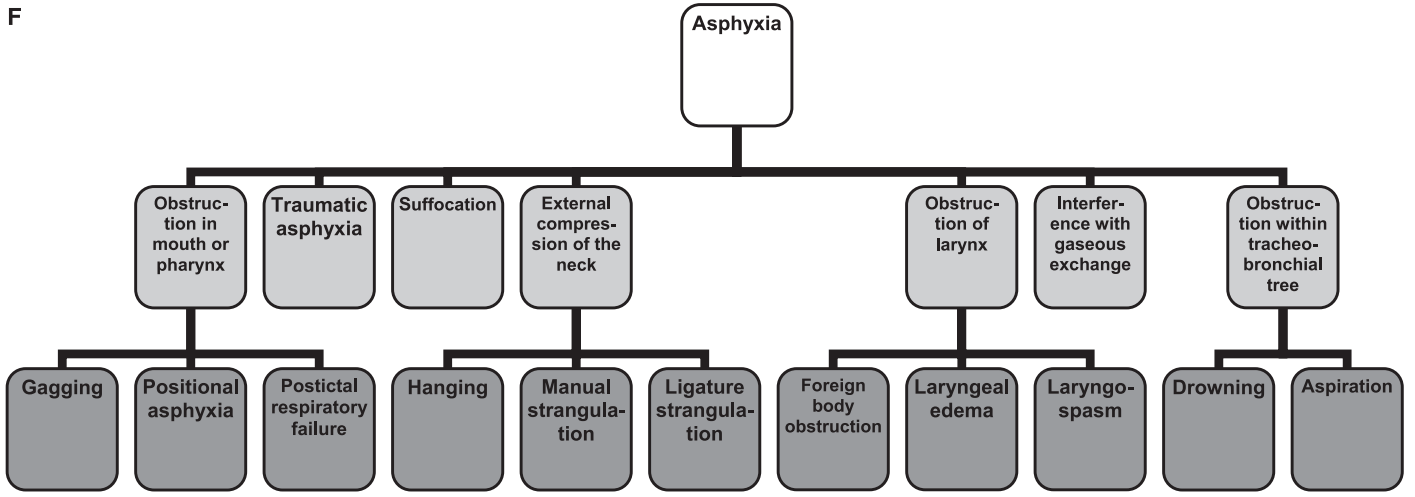


FIG. 1—*Continued*

mechanical means. In keeping with the foregoing authors, the Knight textbook also uses the term mechanical asphyxia in a broader sense, but presents a list of definitions instead of a classification per se (4). Likewise, under the label of ''deaths usually initiated by hypoxic hypoxia or anoxic anoxia'', the textbook by Gordon et al. (8) delineates the different asphyxial types without any attempt of categorization. Finally, the textbook by Spitz (Fig. 1*E*) and the one by Fisher and Petty (Fig. 1*F*) depict two other very distinct classifications of asphyxial deaths (3,9).

Classification and Definition of Types of Asphyxia in the Literature

Suffocation

The term suffocation is not specific. It is a broad term encompassing different types of asphyxia, such as vitiated atmosphere and smothering, associated with a deprivation of oxygen. The majority of authors seem to agree with this definition (Table 1). However, some authors employ this term as a synonym of smothering (1,9,12,13), but this usage is confusing and strongly discouraged. Considering the lack of specificity of this term, it is recommended to avoid using it in certifying death and to replace it by more precise descriptors.

Smothering and Choking

Smothering and choking are both asphyxial deaths caused by an obstruction of the air passages. Depending on the level of the blockage, the condition is called one term or the other. There is no consensus however as to the anatomical landmark serving as a frontier to these entities. For smothering, three different definitions coexist: (i) obstruction at the level of the nose and mouth, (ii) obstruction of the external airways and (iii) obstruction of the upper airways (Table 2*A*). Definitions of choking vary even more widely: (i) synonym of food or foreign body inhalation regardless of the anatomical localization, (ii) obstruction at the level of the mouth, oropharynx and larynx, (iii) obstruction of the larynx, trachea or bronchi, (iv) obstruction of the airways, (v) obstruction of the internal airways, (vi) obstruction of the upper airways and (vii) obstruction of the upper internal airways (Table 2*B*).

TABLE 1—*Definitions of suffocation.*

Reference	Definition
1	Some reserve the term suffocation for deaths in confined spaces, but the terms ''suffocation'' and ''smothering'' are generally used interchangeably
2	Failure of oxygen to reach the blood
4	Not a specific term. Usually refers to a death caused by reduction of the oxygen concentration in the respired atmosphere, formerly called a ''vitiated atmosphere''. Less often used to include smothering or choking
6	Not a specific term. Usually refers to deaths associated with a reduction of available oxygen in respired air. Term often used to include conditions such as smothering
7	Deprivation of oxygen
8	Obstruction to the passage of air into the respiratory tract caused by a closing of the external respiratory orifices. Includes the condition of smothering and overlaying
9	Obstruction at mouth and nose
10	Deprivation of oxygen, either from a lack of oxygen in the surrounding environment or obstruction of the upper airway
11	A broad term encompassing many different types of asphyxia: entrapment, suffocating gases, smothering, choking, mechanical asphyxia, traumatic asphyxia
14	Closure of the nostrils and mouth

TABLE 2—*Definitions of (A) smothering and (B) choking.*

Reference	Definition
(A)	
1	Obstruction at the level of the mouth and nose
2	Asphyxia by mechanical obstruction or occlusion of the external airways, i.e., the nose and mouth
3	Blockage of the nose and mouth
4	Mechanical occlusion of the mouth and nose
6	Blockage of the external air passages
7	Blockage of the external air passage
10	External obstruction of the upper airway
11	A form of suffocation in which the external airways (nose and mouth) are compressed or blocked
15	Obstruction to the external airway, the nose, and the mouth
16	Result from when the nostrils and mouth are covered with the hands or any material which prevents air from entering
(B)	
1	Obstruction at the level of the mouth, oropharynx, larynx
2	Asphyxia by obstruction within the air passages
4	Blockage of the internal airways, usually between the pharynx and the bifurcation of the trachea
6	Obstruction of the upper internal airways usually between the pharynx and main bronchi
7	Blockage of the upper airways by foreign body (bolus, aspiration)
8	Result from the impaction of foreign bodies in the pharynx, larynx, trachea, or bronchi
10	Inhalation of food
11	A form of asphyxia in which the internal airways are obstructed
12	Internal occlusion of the airways, by inhaled foreign bodies or food
14	Result from something impacted in the air passages, usually across the glottis
15	Partial or complete obstruction of the main airway, the larynx, trachea, or bronchi by a foreign body
16	Obstruction of the air passages by a foreign body

Consequently, much confusion is found in the literature's case reports and case series: the localization of the obstruction of air passages does not seem to be very specific to either smothering or choking deaths (Table 3). For example, two very similar asphyxial cases were published in the same year, each reporting death by obstruction of the laryngopharynx by toilet paper in a man with psychiatric problems (17,20). Yet, one team called it smothering while the other considered it as choking. Probably in response to this bewilderment with definitions, some authors do not employ the words smothering or choking and use instead less specific appellations, such as ''aspiration of foreign bodies'' or ''asphyxia by obstruction of the airways'' (29–35).

To propose an anatomical landmark that could be used in the distinction of smothering and choking, it was decided to first look at a less controversial aspect of the problem: the gagging. Gagging is used to muffle screaming usually in the context of robbery with violence. It consists in closing the mouth by applying something over the face or inside the mouth itself, the object sometimes sucked or pushed into the throat (1–4). Most authors regard gagging as a form of smothering (2–4,14,15). In fact, only one author consider it to be strictly choking (11), while another one considers it to be a mixture of smothering and choking (1). Accordingly, the chosen anatomical landmark should classify obstruction in the mouth and throat in the smothering category. Furthermore, the chosen anatomical landmark should be of easy use in the autopsy room. Considering these prerequisites, the epiglottis seems to constitute the ideal landmark: (i) the obstruction of throat and mouth

TABLE 3—*Application of smothering and choking definitions in case reports and case series.*

Reference	Localization of the Obstruction	Classified
17	Laryngopharynx, superior part of the trachea	Smothering
18	Oral cavity, oropharynx, nasopharynx, major airways	Smothering
19	Upper airway: face down in sand	Smothering
	Upper airway: face down in sand	Smothering
	Upper airway: nostrils, mouth, larynx, trachea, bronchi	Choking
	Upper airway: mouth, oropharynx, larynx, trachea	Choking
20	Mouth, oropharynx, laryngopharynx	Choking
21	Throat, palatopharynx, pharynx	Choking
22	Upper airway: upper third of the trachea, larynx	Choking
23	Laryngopharynx	Choking
24	Laryngopharynx, trachea	Choking
25	Tracheal bifurcation, initial portion of the left bronchus	Choking
26	Hypopharynx, larynx	Choking
27	Tracheal bifurcation	Choking
	Back of the mouth extending to the upper trachea	Choking
	Lower part of the trachea, right main bronchus	Choking
28	Epiglottis	Choking

will be classified as smothering; (ii) the identification of an obstruction in relation to the epiglottis is easy at autopsy; (iii) an obstruction located above or below this landmark is unlikely to be moved enough during body manipulation to change the classification of the asphyxia compared to other possible landmarks that are less defined and more open such as the pharynx or larynx. If confronted with an obstruction extending above as well as below the epiglottis, it is recommended to use the lower level of the airway obstruction in classifying the case.

Vitiated Atmosphere, Confined Spaces, and Chemical Asphyxia

The type of asphyxia that occurs when an individual is exposed to an atmosphere depleted in oxygen can be called (i) asphyxia in confined spaces, (ii) entrapment or environmental suffocation, (iii) exclusion of oxygen, (iv) suffocation, (v) death associated with exposure to gases in the atmosphere or (vi) vitiated atmosphere (Table 4). The term suffocation is not recommended, this word being not specific enough. The three labels that appear slightly more common are asphyxia in confined spaces, entrapment, and vitiated atmosphere. Most authors consider this type of asphyxia as a subtype of suffocation (2,4,6,11).

Although there are some slender nuances from one author to the next, for most authors, this category of asphyxia encompasses reduction of oxygen and displacement of oxygen by other gases, as well as gases causing chemical interference with the oxygen uptake and utilization. DiMaio and DiMaio stand alone in this perspective: in their textbook, entrapment and environmental suffocation are considered not only separate entities from suffocating gases and chemical asphyxia, but also separate entities from each other (Table 4). This level of complexity in the subclassification of vitiated atmosphere was not found for any other author. As a matter of fact, all other authors consider that the exclusion of oxygen as well as asphyxia from gases (carbon dioxide, carbon monoxide, methane, and cyanide) are all included in the confined spaces/vitiated atmosphere category (3,4,6,8,10,11). Considering there is no real advantages in complexifying this part of the classification and taking into account the general consensus between authors, it is recommended to join all these forms of asphyxia in a single category, either called confined spaces, entrapment, or vitiated atmosphere.

Mechanical, Postural, Positional, and Traumatic Asphyxia

Mechanical asphyxia has been defined by different authors as either a specific entity characterized by restriction of respiratory movements by external pressure on the chest or abdomen (2,5,7) or as a broad term encompassing several types of asphyxia caused by various mechanical means (1,4,11). This is in keeping with the forensic literature: various authors employ mechanical asphyxia as regrouping certain subtypes of asphyxia, such as hanging, strangulation, throttling, smothering, choking/aspiration, drowning, overlaying, or wedging (36–52). To avoid confusion, it is recommended to keep the phrase mechanical asphyxia as a specific term to designate asphyxia by restriction of respiratory movements.

TABLE 4—*Definitions of asphyxia in confined spaces/entrapment/vitiated atmosphere.*

Reference	Appellation	Definition
1	Confined spaces	An enclosure with limited entry and exit. Has the potential for dangerous atmospheric contamination
2	Entrapment/environmental suffocation	Inadequate oxygen in the environment Entrapment: individuals find themselves trapped in an air-tight or relatively air-tight enclosure; they exhaust the oxygen and asphyxiate Environmental suffocation: an individual inadvertently enters an area where there is gross deficiency of oxygen Excluded suffocating gases and chemical asphyxia
3	Exclusion of oxygen	Because of depletion and replacement by another gas or as a result of chemical interference with its uptake and utilization
4	Suffocation	Death caused by reduction of the oxygen concentration in the respired atmosphere, formerly called vitiated atmosphere; reduction of the oxygen in the atmosphere by physical replacement by other gases or chemical changes such as combustion; by being confined in small airtight space
6	Deaths associated with exposure to gases in the atmosphere	Oxygen may be reduced or absent from respired air or may be displace by the presence of other gases
8	Vitiated atmosphere	A vitiated atmosphere is deficient in oxygen, by displacement of oxygen from the atmosphere by inert gases or by gases generated by the atmosphere
10	Asphyxia with confined and enclosed spaces	Reduced availability of oxygen; nontoxic irrespirable gases
11	Entrapment	A type of suffocation in which an individual is in an airtight or relatively airtight container and gradually consumes the available oxygen until there is no longer enough oxygen to sustain life; entrapment includes gaseous suffocation by gas displacing oxygen, leading to a hypoxic air mixture, and cases in which a substances prevents cells from utilizing oxygen

TABLE 5—*Definitions of positional asphyxia and traumatic asphyxia.*

Reference	
Positional asphyxia (also called postural asphyxia)	
1	A fatal condition owing to the body being oriented in an unusual position, either induced or adopted independently, which mechanically interferes with pulmonary ventilation by airway obstruction and interference of chest wall excursion
2	Individuals become trapped in restricted spaces where, because of the position of their bodies, they cannot move out of that area or position; restriction of their ability to breathe
3	A term to describe situations where the position of an individual interfered with his or her ability to breathe
4	When a person remains in a certain position for an extended time, either because of being trapped, or being in a drunken or drugged state; impedient to adequate respiratory movements
6	A form of smothering, when an individual is incapacitated
11	When an individual acquires a certain body position in which their breathing is compromised, often because of neck twisting with kinking or compression of the trachea and/or elevation of the tongue into the posterior hypopharynx
Traumatic asphyxia	
1	Chest compression by an object weighing more than the victim
2	Occurs when a heavy weight presses down on an individual's chest or abdomen, making respiration impossible. Synonym of mechanical asphyxia
3	Synonym of postural asphyxia and positional asphyxia
4	Mechanical fixation of the chest
6	Interference with the movement of the primary muscles of respiration, namely the intercostal muscles and diaphragm; most often by direct compression by a heavy weight
9	External compression of chest
11	Severe compression of the chest, usually from a large, heavy object

Textbook definitions of positional asphyxia (also called postural asphyxia) and traumatic asphyxia are presented in Table 5. There is general agreement that positional asphyxia refers to a type of asphyxia where the position of an individual compromises the ability to breathe, whereas traumatic asphyxia is defined as a type of asphyxia caused by external chest or abdomen compression by a heavy object. This general agreement is also found in the forensic literature: traumatic asphyxia is related with chest or abdomen compression by a heavy object (53–64) while positional asphyxia is associated with a body position that hinders respiration in all articles (54,65–71) but one (72).

DiMaio and DiMaio classified positional and traumatic asphyxia to be subtypes of suffocation (Fig. 1*A*), as also did the DiMaio and DiMaio's inspired classifications (Fig. 1*B,C*) (2,5,7). However, all other systems classified these two types of asphyxia as nonrelated to suffocation (Fig. 1*D–F*) (1,3,9). As there is no argument to support the inclusion of positional and traumatic asphyxia as subtypes of suffocation, it is recommended to categorize these entities as separate types of asphyxia.

Strangulation and Hanging

Strangulation is a form of asphyxia characterized by closure of the blood vessels and/or air passages of the neck as a result of external pressure on the neck (2,3,6,7,73). All authors define at least two different types of strangulation: ligature strangulation, and manual strangulation. The place of hanging, however, is controversial: several authors consider hanging to be a type of strangulation (2,5,16) or a subtype of ligature strangulation (6,73), whereas other authors consider strangulation and hanging as different entities (3,9–11,14,15).

To determine whether hanging is a form of strangulation or a distinctive separate type of asphyxia is the first problem to solve.

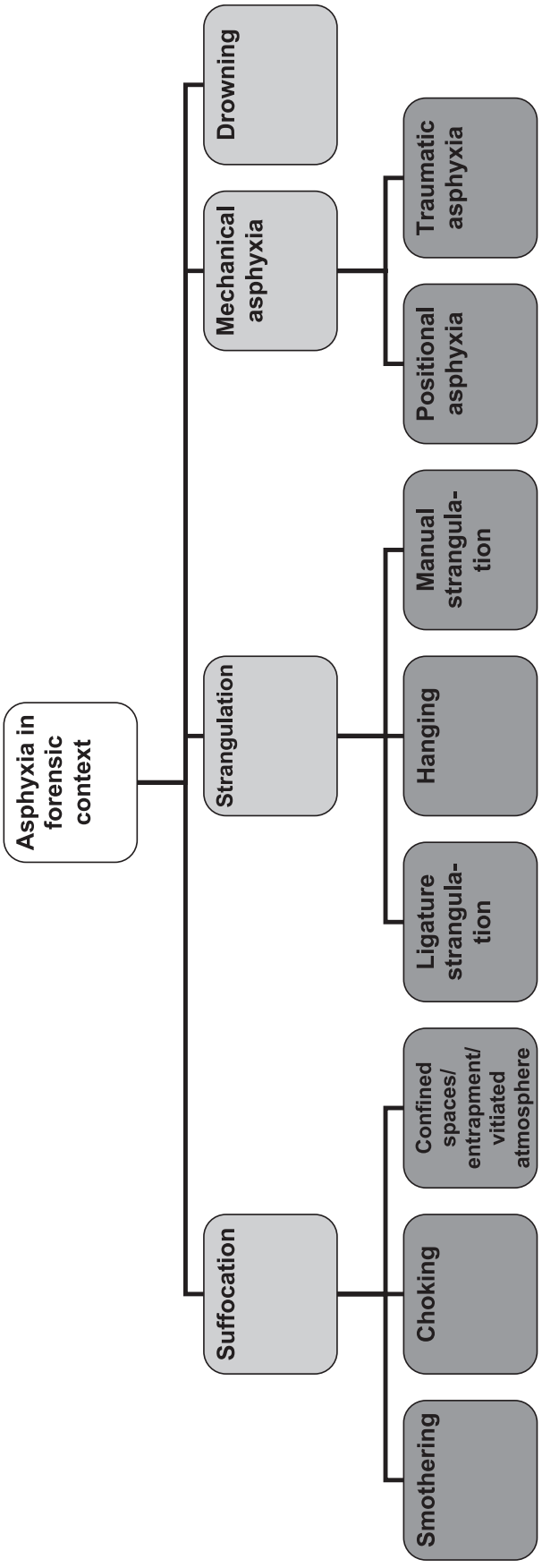


FIG. 2—*The proposed unified classification of asphyxia in forensic context.*

TABLE 6—*Definitions of terms in the proposed unified classification.*

Term	Definition
Suffocation	A broad term encompassing different types of asphyxia such as vitiated atmosphere and smothering, associated with deprivation of oxygen
Smothering	Asphyxia by obstruction of the air passages above the epiglottis, including the nose, mouth and pharynx
Choking	Asphyxia by obstruction of the air passages below the epiglottis
Confined spaces/ entrapment/ vitiated atmosphere	Asphyxia in an inadequate atmosphere by reduction of oxygen, displacement of oxygen by other gases or by gases causing chemical interference with the oxygen uptake and utilization
Strangulation	Asphyxia by closure of the blood vessels and/or air passages of the neck as a result of external pressure on the neck
Ligature strangulation	A form of strangulation in which the pressure on the neck is applied by a constricting band tightened by a force other than the body weight
Hanging	A form of strangulation in which the pressure on the neck is applied by a constricting band tightened by the gravitational weight of the body or part of the body
Manual strangulation	A form of strangulation caused by an external pressure on the structures of the neck by hands, forearms or other limbs
Mechanical asphyxia	Asphyxia by restriction of respiratory movements, either by the position of the body or by external chest compression
Positional or postural asphyxia	A type of asphyxia where the position of an individual compromises the ability to breathe
Traumatic asphyxia	A type of asphyxia caused by external chest compression by a heavy object
Drowning	Asphyxia by immersion in a liquid

Considering that by definition, strangulation is a form of asphyxia by external pressure on the neck, it appears difficult to support the exclusion of hanging from this group.

The second problem is to establish whether hanging is a subtype of ligature strangulation or a separate third type of strangulation. All authors agree that ligature strangulation is defined as a form of strangulation in which the pressure on the neck is applied by a constricting band tightened by a force other than the body weight. All authors also believe that hanging is a form of asphyxia characterized by a constriction of the neck by a ligature tightened by the gravitational weight of the body or part of the body. In fact, the biomechanics of ligature strangulation are identical to those in hanging, except that the tightening force will not be the weight of the body. Nevertheless, these two entities are highly different in their usual settings and all classifications of asphyxia have separated both (Fig. 1*A–F*). Therefore, it is recommended that hanging should be regarded as a type of strangulation, along with manual and ligature strangulation.

It should also be mentioned that two textbook authors reckon that accidental hanging can also occur without a ligature (1,3). For example, they consider accidental death of children by suspension from high chairs, foot of beds, upper edge of car windows or toy boxes as hanging cases. The majority of authors however believe that accidental hangings are restricted to some form of ligature, such as entanglement in ropes, pacifier cord, venetian-blind cords, or drawstrings on clothing (2,6,8,9,15,73). To deem external pressure on the neck by means other than ligature as hanging creates confusion: in the textbook by Spitz (3), these cases are considered accidental hangings in one part of the chapter whereas similar deaths are labeled positional asphyxia later in the same chapter. To avoid such perplexity, it is recommended to restrict the appellation of hanging for cases involving some type of ligature tightened by the weight of the body. Furthermore, it is recommended that all asphyxial deaths caused by external pressure on the neck structures should be labeled strangulation. If a strangulation cannot be subclassified in any of the three types (manual, ligature, or hanging), it should be categorized as strangulation nos (not otherwise specified). For example, a child suspended by the neck by the upper edge of a car window or a hospital patient suspended by the neck by bedrails are cases better classified as strangulation nos.

A special comment on hanging after jumping or being pushed from height (including judicial hanging) is required. This type of hanging is very different in nature from typical hanging, death being related to fracture-dislocation of the upper cervical vertebrae rather than by asphyxia per se. Therefore, it is recommended that this special type of hanging is not included in the classification of asphyxia.

Finally, there is a general agreement that manual strangulation is a form of asphyxia characterized by an external pressure on the structures of the neck by one or both hands, forearms of other limbs.

Additionally, terms such as throttling, garroting, or mugging are occasionally employed. Throttling refers to strangulation, usually by hand or more rarely by ligature (73). Garroting, sometimes used to designate a ligature strangulation, is a former type of Spanish judicial execution, with tightening of a noose around the neck by twisting a rod within the ligature (73). As for mugging, it originally meant the application of pressure to the neck by means of an arm crooked around from the rear, but more recently American usage has widened the term to encompass any kind of robbery with violence (73). These various expressions should be avoided when classifying asphyxial death cases. These terms originate from very specific contexts and by broadening their usage, there is a definite risk of creating more confusion in an already complex and controversial classification.

Drowning

Drowning is defined as death caused by immersion in a liquid, usually water (74–78). The mechanism of such deaths involves a developing hypoxia that ultimately becomes irreversible (74–77). Although asphyxia is certainly an important component of death by drowning, there seems to be a disagreement whether drowning belongs or not in the classification of asphyxia per se. In fact, a few authors leave this entity out from the main classification of forensic asphyxia (Fig. 1*A,E*) while the vast majority includes it in their models (Fig. 1*B–D,F*).

It is recommended therefore that drowning should be included in the forensic classification of asphyxia. However, this inclusion does not necessarily mean that the entity should be discussed in the chapter of asphyxia in textbooks or formal teaching. A better approach would be to include drowning in the classification of asphyxia but discuss it further in the context of investigation of bodies recovered in water.

Classification of Asphyxial Deaths: Proposition of a Unified Model

It is proposed to classify asphyxia in forensic context in four main categories: suffocation, strangulation, mechanical asphyxia, and drowning (Fig. 2). Suffocation subdivides in smothering, choking, and confined spaces/entrapment/vitiated atmosphere. Strangulation includes three separate forms: ligature strangulation, hanging, and manual strangulation. As for mechanical asphyxia, it encompasses positional asphyxia as well as traumatic asphyxia. The definitions of each entity are presented in Table 6.

Conclusion

At this point in time, there is so much variation in the classification and definitions of terms that research and practice are inevitably tinted by confusion. Unfortunately, similar research designs can lead to totally different results depending on the definitions used. Closely comparable cases are called differently by equally competent forensic pathologists. The proposed unified model in this study was designed in an effort to standardize the classification of asphyxia in the forensic context.

References

1. Shkrum MJ, Ramsay DA. Asphyxia. In: Karch SB, series editor. Forensic pathology of trauma: common problems for the pathologist. Totowa, NJ: Humana Press, 2007;65–179.
2. DiMaio VJ, DiMaio D. Asphyxia. In: Geberth VJ, series editor. Forensic pathology, 2nd edn. Boca Raton, FL: CRC Press, 2001;229–77.
3. Spitz WU. Asphyxia. In: Spitz WU, Spitz DJ, editors. Spitz and Fisher's medicolegal investigation of death: guidelines for the application of pathology to crime investigation, 4th edn. Springfield, IL: Charles C Thomas, 2006;783–845.
4. Saukko P, Knight B. Suffocation and 'asphyxia'. In: Ueberberg A, project editor. Knight's forensic pathology, 3rd edn. London, UK: Arnold Publishers, 2004;352–67.
5. Azmak D. Asphyxial deaths: a retrospective study and review of the literature. Am J Forensic Med Pathol 2006;27(2):134–44.
6. Ferris JA. Asphyctic deaths. In: Siegel JA, Saukko PJ, Knupfer GC, editors. Encyclopedia of forensic sciences, vol. 1. London, UK: Academic Press, 2000;308–16.
7. Oehmichen M, Auer RN, König HG. Forensic types of ischemia and asphyxia. In: Oehmichen M, editor. Forensic neuropathology and associated neurology. Berlin: Springer-Verlag, 2005;293–313.
8. Gordon I, Shapiro HA. Deaths usually initiated by hypoxic hypoxia or anoxic anoxia. In: Gordon I, Shapiro HA, editors. Forensic medicine: a guide to principles, 2nd edn. Edinburgh, UK: Churchill Livingstone, 1982;95–129.
9. McNie AB. Asphyxial deaths. In: Fisher RS, Petty CS, editors. Forensic pathology. United Kingdom: Castle House Publications, 1980;123–8.
10. Walker A, Milroy CM, Payne-James J. Asphyxia. In: Payne-James J, Byard RW, Corey TS, Henderson C, editors. Encyclopedia of forensic and legal medicine, vol. 1. Oxford, UK: Elsevier Academic Press, 2005;151–7.
11. Dolinak D, Matshes EW. Asphyxia. In: Dolinak D, Matshes EW, Lew EO, editors. Forensic pathology: principles and practice. Amsterdam: Elsevier Academic Press, 2005;201–24.
12. Nixon JW, Kemp AM, Levene S, Sibert JR. Suffocation, choking and strangulation in childhood in England and Wales: epidemiology and prevention. Arch Dis Child 1995;72:6–10.
13. Simpson K. Asphyxia. In: Simpson K, editor. Forensic medicine, 8th edn. London, UK: Edward Arnold, 1979;91–112.
14. Adams VI, Flomenbaum MA, Hirsch CS. Trauma and disease. In: Spitz WU, Spitz DJ, editors. Spitz and Fisher's medicolegal investigation of death: guidelines for the application of pathology to crime investigation, 4th edn. Springfield, IL: Charles C Thomas, 2006;436–59.
15. Mant KA. Mechanical asphyxia. In: Mant KA, editor. Forensic medicine: observation and interpretation. Chicago, IL: Year Book Publishers, 1960;110–45.
16. Strassmann G, Mass W. Mechanical asphyxia. In: Gradwohl RBH, editor. Legal medicine. St-Louis, MO: The C.V. Mosby Company, 1954;260–84.
17. Saint-Martin P, Bouyssy M, O'Byrne P. An unusual case of suicidal asphyxia by smothering. J Forensic Leg Med 2007;14(1):39–41.
18. Burke MP, Path DF, Alamad S, Dip G, Opeskin K. Death by smothering following forced quetiapine administration in an infant. Am J Forensic Med Pathol 2004;25(3):243–5.
19. Hanson KA, Gilbert JD, James RA, Byard RW. Upper airway occlusion by soil—an unusual cause of death in vehicle accidents. J Clin Forensic Med 2002;9(2):96–9.
20. Sauvageau A, Yesovitch R. Choking on toilet paper: an unusual case of suicide and a review of the literature on suicide by smothering, strangulation, and choking. Am J Forensic Med Pathol 2006;27(2):173–4.
21. Fernando GC. A case of fatal suffocation during an attempt to swallow a pool ball. Med Sci Law 1989;29(4):308–10.
22. Pinheiro J, Cordeiro C, Vieira DN. Choking death on a live fish (Dicologoglossa cuneata). Am J Forensic Med Pathol 2003;24(2):177–8.
23. Kurihara K, Kuroda N, Murai T, Shinozuka T, Yanagida J, Matsuo Y, et al. A case of homicidal choking mistaken for suicide. Med Sci Law 1992;32(1):65–7.
24. Murty OP, Mun K, Gopinath N, Wong KT. Choking on food: a rare case of alexander leukodystrophy and choking. Am J Forensic Med Pathol 2008;29(4):364–7.
25. Pampin JB, Varela LG. Suicidal choking caused by a bizarre combination of inhalation to the bronchi and external neck compression. Leg Med (Tokyo) 2001;3(2):119–22.
26. Kumar MV, Venkatesh VT, Jagannatha SR. Fast eating syndrome: a case report. Med Sci Law 2008;48(1):78–81.
27. Seymour A, Black M, McFarlane JH, Oliver JS. Death by obstruction: sudden death resulting from impromptu ingestion of drugs. Am J Forensic Med Pathol 2003;24(1):17–21.
28. Meel BL. An accidental suffocation by a rubber balloon. Med Sci Law 1998;38(1):81–2.
29. Adelman HC. Asphyxial deaths as a result of aspiration of dental appliances: a report of three cases. J Forensic Sci 1988;33(2):389–95.
30. Bhana BD, Gunaselvam JG, Dada MA. Mechanical airway obstruction caused by accidental aspiration of part of a ballpoint pen. Am J Forensic Med Pathol 2000;21(4):362–5.
31. Kettner M, Ramsthaler F, Horlebein B, Schmidt PH. Fatal outcome of a sand aspiration. Int J Legal Med 2008;122(6):499–502.
32. Szentmariay IF, Laszik A, Sotonyi P. Sudden suffocation by surgical sponge retained after a 23-year-old thoracic surgery. Am J Forensic Med Pathol 2004;25(4):324–6.
33. Collins KA, Presnell SE. Asphyxia by tracheobronchial thrombus. Am J Forensic Med Pathol 2005;26(4):327–9.
34. Njau SN. Adult sudden death caused by aspiration of chewing gum. Forensic Sci Int 2004;3:103–6.
35. Colombage SM. Laryngeal obstruction by heroin packets. Am J Forensic Med Pathol 2003;24(2):153–4.
36. Zhu BL, Ishida K, Quan L, Taniguchi M, Oritani S, Li DR, et al. Postmortem serum uric acid and creatinine levels in relation to the causes of death. Forensic Sci Int 2002;125(1):59–66.
37. Zhu BL, Ishida K, Fujita MQ, Maeda H. Immunohistochemical investigation of a pulmonary surfactant in fatal mechanical asphyxia. Int J Legal Med 2000;113(5):268–71.
38. Kohli A, Verma SK, Agarwal BB. Accidental strangulation in a rickshaw. Forensic Sci Int 1996;78(1):7–11.
39. Tamaki K, Sato K, Katsumata Y. Enzyme-linked immunosorbent assay for determination of plasma thyroglobulin and its application to postmortem diagnosis of mechanical asphyxia. Forensic Sci Int 1987;33(4):259–65.
40. Brinkmann B, Fechner G, Püschel K. Identification of mechanical asphyxiation in cases of attempted masking of the homicide. Forensic Sci Int 1984;26(4):235–45.
41. Verma SK, Aggarwal NK, Kohli A. Accidental ligature strangulation deaths in East Delhi (India). Med Sci Law 2005;45(1):47–51.
42. Dada MA. Laryngeal cyst and sudden death. Med Sci Law 1995;35(1):72–4.
43. Quan L, Zhu BL, Ishida K, Oritani S, Taniguchi M, Fujita MQ, et al. Intranuclear ubiquitin immunoreactivity of the pigmented neurons of the substantia nigra in fatal acute mechanical asphyxiation and drowning. Int J Legal Med 2001;115(1):6–11.
44. Katsumata Y, Sato K, Oya M, Yada S. Detection of thyroglobulin in blood stains as an aid in the diagnosis of mechanical asphyxia. J Forensic Sci 1984;29(1):299–302.

45. Deming JE, Mittleman RE, Wetli CV. Forensic science aspects of fatal sexual assaults on women. J Forensic Sci 1983;28(3):572–6.
46. Shetty M, Shetty BS. Accidental ligature strangulation due to electric grinder. J Clin Forensic Med 2006;13(3):148–50.
47. Turillazzi E, D'Errico S, Neri M, Fineschi V. An unusual mechanical asphyxia in a homicide-suicide case by smothering and strangulation. Am J Forensic Med Pathol 2006;27(2):166–8.
48. Di Nunno N, Vacca M, Costantinides F, Di Nunno C. Death following atypical compression of the neck. Am J Forensic Med Pathol 2003;24(4): 364–8.
49. Collins KA. Death by overlaying and wedging: a 15-year retrospective study. Am J Forensic Med Pathol 2001;22(2):155–9.
50. Kogan Y, Bloom T. Suicidal ligature strangulation with an elastic band. Am J Forensic Med Pathol 1990;11(4):329–30.
51. Rao VJ, Wetli CV. The forensic significance of conjunctival petechiae. Am J Forensic Med Pathol 1988;9(1):32–4.
52. Vieira DN, Pinto AE, Sá FO. Homicidal hanging. Am J Forensic Med Pathol 1988;9(4):287–9.
53. Byard RW, Wick R, Simpson E, Gilbert JD. The pathological features and circumstances of death of lethal crush/traumatic asphyxia in adults—a 25-year study. Forensic Sci Int 2006;3:200–5.
54. Wankhede AG, Dongre AP. Head injury with traumatic and postural asphyxia: a case report. Med Sci Law 2002;42(4):358–9.
55. Miyaishi S, Yoshitome K, Yamamoto Y, Naka T, Ishizu H. Negligent homicide by traumatic asphyxia. Int J Legal Med 2004;118(2):106–10.
56. Betz P, Beier G, Eisenmenger W. Pulmonary giant cells and traumatic asphyxia. Int J Leg Med 1994;106(5):258–61.
57. Hitchcock A, Start RD. Fatal traumatic asphyxia in middle-aged man in association with entrapment associated hypoxyphilia. J Clin Forensic Med 2005;12(6):320–5.
58. Byard RW. The brassiere 'sign'—a distinctive marker in crush asphyxia. J Clin Forensic Med 2005;12(6):316–9.
59. Gill JR, Landi K. Traumatic asphyxial deaths due to an uncontrolled crowd. Am J Forensic Med Pathol 2004;25(4):358–61.
60. Lau G. Pulmonary cartilage embolism: fact or artefact? Am J Forensic Med Pathol 1995;16(1):51–3.
61. Nichols GR 2nd, Davis GJ, Parola AC. Dirty diving. Sudden death of a scuba diver in a water treatment facility. Am J Forensic Med Pathol 1992;13(1):72–5.
62. Wolodzko AA, Taff ML, Ratanaproeska O, Spitz WU. An unusual case of compression asphyxia and smothering. Am J Forensic Med Pathol 1986;7(4):354–5.
63. Kohr RM. Inflicted compressional asphyxia of a child. J Forensic Sci 2003;48(5):1148–50.
64. Ely SF, Hirsch CS. Asphyxial deaths and petechiae: a review. J Forensic Sci 2000;45(6):1274–7.
65. Padosch SA, Schmidt PH, Kröner LU, Madea B. Death due to positional asphyxia under severe alcoholisation: pathophysiologic and forensic considerations. Forensic Sci Int 2005;149(1):67–73.
66. Glatter K, Karch SB. Positional asphyxia: inadequate oxygen, or inadequate theory? Forensic Sci Int 2004;3:201–2.
67. Byard RW. Hazardous infant and early childhood sleeping environments and death scene examination. J Clin Forensic Med 1996;3(3): 115–22.
68. Busuttil A, Obafunwa JO. Recreational abdominal suspension: a fatal practice. A case report. Am J Forensic Med Pathol 1993;14(2):141–4.
69. Reay DT, Fligner CL, Stilwell AD, Arnold J. Positional asphyxia during law enforcement transport. Am J Forensic Med Pathol 1992;13(2): 90–7.
70. Bell MD, Rao VJ, Wetli CV, Rodriguez RN. Positional asphyxiation in adults. A series of 30 cases from the Dade and Broward County Florida Medical Examiner Offices from 1982 to 1990. Am J Forensic Med Pathol 1992;13(2):101–7.
71. Belviso M, De Donno A, Vitale L, Introna F. Positional asphyxia. Reflection on 2 cases. Am J Forensic Med Pathol 2003;24(3):292–7.
72. O'Halloran RL, Dietz PE. Autoerotic fatalities with power hydraulics. J Forensic Sci 1993;38(2):359–64.
73. Saukko P, Knight B. Fatal pressure on the neck. In: Ueberberg A, project editor. Knight's forensic pathology, 3rd edn. London, UK: Arnold Publishers, 2004;368–94.
74. Shkrum MJ, Ramsay DA. Bodies recovered in water. In: Karch SB, series editor. Forensic pathology of trauma: common problems for the pathologist. Totowa, NJ: Humana Press, 2007;243–93.
75. DiMaio VJ, DiMaio D. Death by drowning. In: Geberth VJ, series editor. Forensic patholog, 2nd edn. Boca Raton, FL: CRC Press, 2001; 399–407.
76. Spitz DJ. Investigation of bodies in water. In: Spitz WU, Spitz DJ, editors. Spitz and Fisher's medicolegal investigation of death: guidelines for the application of pathology to crime investigation, 4th edn. Springfield, IL: Charles C Thomas, 2006;846–81.
77. Saukko P, Knight B. Immersion deaths. In: Ueberberg A, project editor. Knight's forensic pathology, 3rd edn. London, UK: Arnold Publishers, 2004;395–411.
78. Bell MD. Drowning. In: Dolinak D, Matshes EW, Lew EO, editors. Forensic pathology: principles and practice. Amsterdam: Elsevier Academic Press, 2005;227–37.

Additional information and reprint requests:
Anny Sauvageau, M.D., M.Sc.
Office of the Chief Medical Examiner
7007, 116 Street
Edmonton, AB T6H 5R8
Canada
E-mail: anny.sauvageau@gmail.com

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01435.x
Available online at: interscience.wiley.com

PAPER

PATHOLOGY AND BIOLOGY

Renaud Clément,[1,2] *M.D., M.Sc.; Margaret Redpath,*[1] *M.D.; and Anny Sauvageau,*[1] *M.D., M.Sc.*

Mechanism of Death in Hanging: A Historical Review of the Evolution of Pathophysiological Hypotheses

ABSTRACT: In cases of hanging, the exact mechanism leading to death has yet to be elucidated. Most of our contemporary knowledge is still based on writings from the end of the 19th and the beginning of the 20th century. This article reviews the historic experiments that shaped our current theories. Medico-legal textbooks written in English and French from 1870 to 1930 were reviewed. Various animals, such as rabbits, mice, and dogs, have been used to develop animal models of hanging. Limited human studies on cadavers and judicial hangings have provided some additional insight into the pathophysiology of death by hanging. The main pathophysiological theories described were respiratory asphyxia, interruption to cerebral blood flow because of occlusion of vessels in the neck, and cardiac inhibition secondary to nerve stimulation. The relative contributions of each of these theories to death in cases of hanging is still debated today.

KEYWORDS: forensic science, hanging, asphyxia, pathophysiology, historic, respiratory asphyxia, vascular occlusion, cardiac inhibition

Recently, filmed hangings have been used as a powerful tool in understanding the pathophysiology of human asphyxia (1–3). The Working Group on Human Asphyxia (WGHA) was formed in 2006, and since its creation, 8 filmed hangings have been analyzed. Observing the videos reveals that loss of consciousness occurs quickly, followed by convulsions and a complex pattern of alternating phases of decerebrate and decorticate rigidity. The videos also demonstrate some auditory evidence of persistent air passage through the airways during the hanging process.

The study of the WGHA has provided interesting new insight into the pathophysiology of asphyxia by hanging. Before these new developments, most of our contemporary knowledge was based on writings from the end of the 19th and the beginning of the 20th century. In this article, the literature from this crucial turning point in the construction of modern views on asphyxia by hanging will be reviewed (period of 1870–1930). It is important to understand the origin of our current theories and the models that were used to develop them so that we can knowledgably reevaluate their validity.

Materials and Methods

A search for medico-legal textbooks written in English and French was carried out. All available textbooks from the 60-year study period (1870–1930) containing a chapter on "hanging" were selected. Using these selection criteria, ten textbooks were eligible for the study (4–13). The reading focused on the pathophysiological hypothesis accepted or rejected by the authors: occlusion of the trachea, occlusion of vessels, and pneumogastric nerve stimulation. Experiments supporting the authors' positions were compiled, as well as their clinical observations.

Results and Discussion

At the beginning of the study period, in France in 1870, Tardieu (4) was convinced that the mechanism of death in hanging was respiratory asphyxia, i.e., an asphyxia caused by occlusion of the trachea preventing lung ventilation and gas exchange. Based on the knowledge available at the time, Tardieu concluded that hanging was not associated with cerebral ischemia because intimal tears of the carotid and congestion of the brain were uncommonly encountered at autopsy. In the following decade, an Austrian author, Hoffman, shared this view that hanging was a form of respiratory asphyxia, but he questioned the proposed mechanism. He believed that the cartilaginous structures of the larynx and trachea could not be sufficiently compressed to cause obstruction. Instead, he proposed that it was the compression of these structures against the posterior pharyngeal wall that caused the obstruction, sometimes in association with occlusion by the base of the tongue (5). Hoffman repeatedly observed that he was not able to infuse the neck vessels of hanging victims, including those of children. He concluded that the external pressure of the neck ligature was superior to the blood pressure, which led him to propose that occlusion of the neck vessels also occurs during the hanging process.

In 1893, in England, Dixon (6) further emphasized the crucial role of respiratory asphyxia in death by hanging. In support of his theory, he cited the work of Langreuter (14), a German researcher. By tying a rope around the neck of cadavers who had died of

[1]Laboratoire de sciences judiciaires et de médecine légale, Edifice Wilfrid-Derome, 1701, Parthenais street, 12th floor, Montreal, QC H2K 3S7, Canada.
[2]Laboratoire de médecine légale, Faculty of Medicine, Nantes, 1 Gaston Veil street, 44035 Nantes Cedex 1 France.

Received 31 Mar. 2009; and in revised form 8 July 2009; accepted 8 August 2009.

natural causes, Langreuter was able to directly visualize the effects of neck compression on the laryngotracheal area. He demonstrated that with moderate neck compression, the epiglottis becomes crushed against the posterior pharyngeal wall and obstructs the upper airways. With a more intense compression, not only the epiglottis but also the base of the tongue was pushed back against the posterior pharyngeal wall. Apart from this interesting autopsy study, Dixon (6) also substantiated his point of view by reporting the case of a colleague, Ecker, in which the frozen body of a hanging man was found and similar anatomic evidence of compression was observed at autopsy.

Although Dixon favored respiratory asphyxia as the main pathophysiological factor leading to death in hanging, he also considered the possibility that the noose causes cervical constriction and obstruction of the vessels of the neck. He reviewed experiments with tracheotomized rabbits that were performed by Reineboth (1895). This author found that even if the noose was tightened above the level of the tracheotomy opening, the rabbits still died. The rabbits survived longer when the cervical constriction was placed above the opening compared to below, but they still died (5–6 min vs. 15–20 min). This medico-legal discovery contradicted two concurrent observations that were made using dogs. The first observation was that a dog could be resuscitated after 3 min of hanging when a tracheotomy was performed. It was also shown that tracheotomized dogs that were hanged did not die if the level of compression was above the opening for air circulation. In fact, when the animal was hanged with incomplete occlusion of the airways, the length of survival was increased; thus the mechanism of death was believed to be a lack of intracranial circulation. It should be mentioned, however, that this discrepancy between the rabbit and the dog models is not surprising considering contemporary knowledge of their circulatory systems; in rabbits, both the internal carotids and the vertebral arteries contribute to cerebral circulation, whereas in dogs, the internal carotids are relatively less developed and the vertebral arteries provide the majority of the blood flow to the brain (15). Furthermore, dogs have highly developed anastomoses from the external carotids that contribute to the collateral blood supply of the brain. Therefore, it was recently proposed to avoid using dogs as an animal model of hanging (15).

Around the same time as Dixon (1895), two other English authors did not accept respiratory asphyxia as the main pathophysiological explanation of death in all cases of hanging (6). Although Guy and Ferrier (7) thought that respiratory asphyxia was implicated in deaths caused by complete hanging, they believed that occlusion of blood flow plays a more important role in deaths caused by partial hanging. At the end of the 19th century, Brouardel (8) confirmed the hypothesis that the mechanism of death in hanging is the tightening of a cervical noose that causes occlusion of the vasculature of the neck. He observed the level of anemia in the retinas of hanged animals, which is correlated with cerebral ischemia. The length of time that it took to die from hanging was compared between dogs that were and were not tracheotomized. Death occurred, respectively, in 15–20 and 6 min.

Another experiment by Brouardel (8) demonstrated that the jugular vein is occluded by a pressure of 2 kg, the carotid artery by a pressure of 5 kg, and the vertebral artery by a pressure of 30 kg. These occlusion pressures were estimated by experimenting with dead bodies. To gain access to the intracranial vessels, the skull was opened and the brain was removed. Then, a ligature was positioned around the neck and a dynamometer was used to measure the force on the ligature. Access to the vessels below the level of the ligature was prepared for water perfusion. The body was then partially lifted from the table by the ligature, thus creating a human model of incomplete hanging. The weight on the dynamometer was correlated with the degree of occlusion of the vessels that was assessed by infusing water. To assess the weight of occlusion for the airways, a tracheotomy was performed. Brouardel stated in his book that by 15 kg of weight on the neck, the pressure of the base of the tongue on the pharynx is sufficient to completely block airflow. To further support this opinion, he described large retropharyngeal ecchymoses in the victims of judicial hangings.

Lacassagne (9) felt that there was a simple, clinical distinction to be made about the mechanism of death in hanging: a complete hanging causes respiratory obstruction, whereas an incomplete hanging causes compression of the vasculature of the neck. Lacassagne confirmed that only 15 kg of pressure on the trachea is required to cause anatomical changes to the soft tissue of the larynx severe enough to cause respiratory occlusion. Derome (10) questioned the implication of respiratory asphyxia as a mechanism of death after reviewing the systematic hangings of tracheotomized rabbits. Balthazard (11), Smith (12), and Webster (13) continued to publish that the principal mechanism of death in hanging was respiratory asphyxia and that vascular occlusion was a minor contributor.

Hofmann (5) and Brouardel (8) considered the possibility of a different mechanism of death in hanging: cardiac inhibition caused by the stimulation of pericarotid nerves. Hofmann (5) believed that bilateral vagal stimulation could provoke cardiac arrest. Lacassagne (9), Balthazard (11), and especially Smith (12) and Webster (13) accepted cardiac inhibition as a possible mechanism of death in hangings. Derome (10) accepted the theory in cases where the nerves were torn. Death because of the inhibitory effect of stimulation of the pericarotid nerves during hanging was a theory that was rejected by Tardieu (4), Dixon (6), and Guy and Ferrier (7). For these authors, experimentation and observation clearly showed the persistence of cardiac activity after the total cessation of body movement and respiration. The pathophysiological theories of the mechanism of death in hanging, held by the different authors mentioned in this text, are summarized in Table 1.

Loss of consciousness, convulsions, and apparent death were the classic clinical signs described by all authors. The time of onset of these signs was generally vague or was omitted (Table 2). Hofmann (5) used the information he gathered from assisting judicial hangings and talking to people who had survived hangings to describe these three phases in detail. In addition to these three phases, he also described that prior to losing consciousness, hanging victims experience intense feelings of pleasure and well-being. Forty years later, Balthazard (11) provided a more precise estimation of the timing of the phases.

Considering how quickly loss of consciousness occurs, Tardieu (4) and Balthazard (11) thought that it could only be induced by vagal stimulation. This issue was intensely debated because other authors claimed the opposite. According to them, the rapidity of the onset of the loss of consciousness ruled out vagal stimulation as the main mechanism and was more supportive of airway or vessel obstruction (4,5,8).

Because of the early onset of loss of consciousness, some believed that victims did not suffer. The convulsive fits were thought to be related to hypercapnia (8). Only Hofmann (5) described convulsive movements lasting for a 30-sec time period. No mention was made of signs of decerebrate or decorticate rigidity.

The delay to the onset of death was thought to differ based on the pathophysiologial mechanism involved (8). According to Brouardel, rapid death within 5–10 min indicated respiratory asphyxia, slower death lasting 12–20 min suggested occlusion of

TABLE 1—*Summary of the proposed pathophysiology of hanging by author.*

Authors (Ref)	Respiratory Asphyxia	Vascular Occlusion	Vagal Irritation	Observation (Obs)/ Experimentation (Exp)
Tardieu (4)	Yes, MM	No	No	Obs: hanging cases
Hofmann (5)	Yes, MM	Yes, MM	Yes, SM	Obs: judicial hangings Exp: cadaveric carotid perfusion
Dixon (6)	Yes, MM	Yes, PM	No	Exp: delayed survival time of hanging rabbits Obs: reviewed the work of Langreuter (reporting the effects on neck tissues of tying a rope) and Ecker (sections of frozen hanging man).
Guy and Ferrier (7)	Yes, MM	Yes, PM	Refuted	Obs: cardiac activity persists after death.
Brouardel (8)	Yes, MM	Yes, PM	Yes, SM	Exp: study of pressure required to occlude the vasculature of the neck
Lacassagne (9)	Yes	Yes	No	Reproduced Brouardel's experiments
Derome (10)	Undecided	Yes, PM only arterial blockage	Yes, SM	Obs: study of hanging tracheotomized rabbits
Balthazard (11)	Yes	Yes	Yes, SM	Similar to previous data
Smith (12)	Yes, MM	Yes, PM	Yes, SM	Obs
Webster (13)	Yes, MM	Yes, PM	Yes, SM	No obs. or exp

MM, main mechanism; PM, possible mechanism; SM, suspected mechanism.

TABLE 2—*Summary of the clinical descriptions of hanging provided by the authors.*

Author(s) (Ref)	Onset of Loss of Consciousness	Onset of Convulsions	Apparent Death from Respiratory Asphyxia
Tardieu (4)	Fast	½ min	5–10 min
Hofmann (5)	Immediately after hanging (observations of judicial hangings and people resuscitated after hanging)	1 min after hanging and during 30 sec	Not reported
Brouardel (8)	<1 min	Not reported	7–8 min/12–20 min (Time delay associated with occlusion of vasculature of the neck)
Lacassagne (9)	5–6 sec arterial obstruction (in cases of anterior noose) 8–9 sec venous blockade (in cases of posterior noose)	Correlated with hypercapnia	5–10 min
Balthazard (11)	½ min. Later in cases of vascular occlusion	2 min	10 min
Dixon, Guy and Ferrier, Derome, Smith, Webster) (6,7,10,12,13)	Not reported	Not reported	Not reported

the vasculature of the neck, and an even slower death lasting 15–20 min implied cardiac inhibition.

Only one author provided a detailed description of the absence and presence of respiratory movements during hanging. Hofmann (5) described a complex sequence of alternating apnea, dyspnea, and superficial respiratory movements. After an initial apnea lasting 1 min, the victim began having inspiratory dyspnea, followed by convulsions, then after 30 sec, a deep expiratory movement that led to a second phase of apnea that ended after 1 min with a deep inspiration. A third phase of apnea lasting 1 min passed, and 5–10 shallow inspiratory and expiratory movements occurred over 1–2 min.

Conclusion

Data collected by these eminent scientists from the end of the 19th century and the beginning of the 20th century are still of interest today. The vast majority of these authors described the rapid loss of consciousness also observed in contemporary studies. Their description of the subsequent body responses to asphyxia (convulsions, decerebrate rigidity, decorticate rigidity) is not as detailed, however, as the contemporary studies on filmed hangings (2,3).

This historical review supports the fundamental scientific principle that just because a theory is accepted does not guarantee that it reflects the truth. In hanging deaths, respiratory asphyxia by occlusion of the airway was considered the principal mechanism of death for several decades before the theories of vascular occlusion and cardiac inhibition were gradually accepted.

Despite a long history of being investigated, the pathophysiology of hanging still needs to be revisited and studied in the 21st century, as the relative weight of the main three mechanisms of death in hanging is still debated. The systematic study of filmed human hangings and the development of better animal models will someday provide a definite answer to this old query.

Acknowledgments

The authors thank Dr. Geoffroy de la Grandmaison, a forensic pathologist in Garches, France, for providing copies of several French textbooks.

References

1. Sauvageau A, Racette S. Agonal sequences in a filmed suicidal hanging: analysis of respiratory and movement responses to asphyxia by hanging. J Forensic Sci 2007;52(4):957–9.
2. Sauvageau A. Agonal sequences in four filmed hangings: analysis of respiratory and movement responses to asphyxia by hanging. J Forensic Sci 2009;54(1):192–4.
3. Sauvageau A, LaHarpe R, Geberth VJ. Agonal sequences in eight filmed hangings: analysis of respiratory and movement responses to asphyxia by hanging. J Forensic Sci. E-pub ahead of print. DOI: 10.1111/j.1556-4029.2010.01434.x.
4. Tardieu A. Étude médico-légale sur la pendaison. Ann Hyg Pub Med Lég 1870;33:78–128.
5. Hofmann E. Nouveaux éléments en médecine légale. Paris: JB Baillière, 1881.
6. Dixon JC. Forensic medicine and toxicology. London: Griffin & Co editors, 1893.
7. Guy W, Ferrier D. Principles of forensic medicine. 7th rev. London: William & R, 1895.
8. Brouardel P. La pendaison, la strangulation, la suffocation, la submersion. Paris: J.B Bailliere et fils, 1897.

9. Lacassagne A. Précis de médecine légale. Paris: Masson, 1906.
10. Derome W. Précis de médecine légale. Montreal: Compagnie d'imprimerie des Marchands Ltée, 1920.
11. Balthazard V. Précis de médecine légale. Paris: J-B Baillière, 1921.
12. Smith S. A text-book for students and practitioners. Forensic medicine. London: J. & A. Churchill Ltd, 1925.
13. Webster R. Legal medicine and toxicology. Philadelphia and London: WB Saunders, 1930.
14. Langreuter J. Über die mechanischen verhältnisse des strangulationstodes. Vjschr Ger Med 1886;XLV:295–309.
15. Boghossian E, Clément R, Redpath M, Sauvageau A. Respiratory, circulatory and neurological responses to asphyxia by hanging: a review of animal models. J Forensic Sci. E-pub ahead of print. DOI: 10.1111/j.1556-4029.2010.01436.x.

Additional information and reprint requests:
Anny Sauvageau, M.D., M.Sc.
Office of the Chief Medical Examiner
7007, 116 Street
Edmonton, AB T6H 5R8
Canada
E-mail: anny.sauvageau@gmail.com

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01436.x
Available online at: interscience.wiley.com

PAPER

PATHOLOGY AND BIOLOGY

Elie Boghossian,[1] *B.Sc.; Renaud Clément,*[1,2] *M.D.; Margaret Redpath,*[1] *M.D.; and Anny Sauvageau,*[1] *M.D., M.Sc.*

Respiratory, Circulatory, and Neurological Responses to Hanging: A Review of Animal Models

ABSTRACT: The pathophysiology of hanging is still poorly understood. This article presents a review of eight animal models: four models of isolated occlusion of the vessels of the neck (group 1), one model of combined tracheal and vessel occlusion (group 2), and three models of true animal hanging (group 3). Occlusion of the airway passages in group 2 did not accelerate respiratory arrest compared to group 1. Cessation of cerebral blood flow, rather than airway obstruction, seems to be the main cause of respiratory decline. In general, muscular movements ceased after 1–3.5 min and early generalized tonic-clonic convulsions were described. Complete circulatory collapse seems to occur between 4 and 8.5 min. These observations from animal models of hanging are compared with the data collected from filmed human hangings. Avenues to improve animal models are discussed.

KEYWORDS: forensic science, asphyxia, hanging, animals, rats, pathophysiology

Death by hanging is induced by obstruction of the blood vessels and/or airways of the neck by a ligature that is tightened by the victim's own body weight (1). The exact pathophysiology of hanging, however, is still poorly understood. Despite great advances in forensic science over the last few decades, several questions related to hanging remain unanswered. How long does it take to lose consciousness? How long does it take to have irreversible brain damage? How does cerebral ischemia specifically cause cardiopulmonary arrest?

For ethical reasons, it is not feasible to conduct studies about hanging on human subjects. An indirect way to study human hangings was introduced with the analysis of films of autoerotic or, less commonly, suicidal nature (2,3). Animal models can also substitute for human experimentation. In fact, animal studies simulating hanging have been carried out, but the applicability of the results to humans remains unknown. This article presents a review of animal studies evaluating hanging and assesses the respiratory, circulatory, and neurological responses observed in the various animal models. Moreover, physiological responses to hanging observed in these animal models are compared to the physiological responses described in humans.

Review of Animal Studies

The 20th century literature was reviewed for all animal models reproducing some aspect of hanging. Medline search and cross-reference search was used to identify studies, as well as cross-referencing from major forensic textbooks. A total of eight studies were found (4–11). The animal models included a variety of different species (cats, dogs, rabbits, and rats), and the methods used by the authors varied greatly. These studies can be broadly divided into three different groups.

Animal Studies: Group 1

In the earliest group, from 1930 to 1948, cerebral blood flow was interrupted in cats and dogs by ligaturing or clamping the carotid and/or the vertebral arteries to explore the effects of cerebral ischemia both immediately and during the recovery phase (Table 1). These studies were not intended to be used to analyze the pathophysiology of hanging; consequently, the animal model lacked the concomitant airway occlusion that occurs in the classic conception of hanging.

Gildea et al. (4) used young cats as an animal model to study cerebral anemia. While the animal was under light ether anesthesia, the vertebral vessels of the neck were dissected out and cord ligatures were placed around them that could later be tightened to occlude blood flow to the brain. The animals were artificially ventilated by way of a tracheal tube. The degree of cerebral anemia was estimated to be severe with the presence of two symptoms: respiratory arrest within 1–2 min of occlusion of the vessels and the presence of convulsions during occlusion of the vessels. The brain was later evaluated histologically.

Sugar et al. (5) also used cats as their model. While the cats were under Nembutal anesthesia, their vertebral arteries were ligated and lifting ligatures were placed on the carotids. One side of the calvarium was removed to allow for the placement of the electrodes, and the contralateral carotid was then occluded while

[1]Laboratoire de sciences judiciaires et de médecine légale, Montreal, Canada.
[2]Laboratoire de médecine légale, Faculty of Medicine, Nantes, 1 Gaston Veil street, 44035 Nantes Cedex 1 France.

Received 31 Mar. 2009; and in revised form 8 July 2009; accepted 8 Aug. 2009.

TABLE 1—*The course of respiratory, circulatory, and neurological responses following animal asphyxia simulating hanging with obstruction of the blood vessels in the neck.*

Reference	Model	Objective	Methods	Respiratory Component	Neurological Component	Circulatory Component
(4)	Cat (n = 90)	To reproduce experimentally lesions of the cerebral cortex associated with ischemia, to study the tolerance of cerebral nerve cells to ischemia and to correlate these lesions with the symptoms	• Ligature of carotid and vertebral arteries • Airways open (tracheal intubation)	Loss of respiratory movements in 1–2 min (preceded by irregular and gasping respiratory movements)	• Animals stiffened out immediately, became limp and reflexes disappeared • Convulsions (in 65/90 cats): sometimes preceded by clonic movements of one foot or rigidity of the whole body, followed by a brief period of tetanic convulsions and then gross clonic convulsions of the whole body	• Immediate increase of pulse rate • The heart slows down with cessation of respiration, then briefly becomes rapid again between 2 and 4 min later (with failure of vagus center)
(5)	Cat (n = 15)	To study the effect of ischemia on brain electric potentials in different cephalic regions and to evaluate survival and recovery times	• Ligature of carotid and vertebral arteries (during 20–180 sec) • Airways open (tracheal intubation)	Immediate increase in respiratory movements followed by a progressive decrease, ending in apnea in 30–40 sec	• General tonic and clonic movements observed from 20th to 60th sec • Passive posture and flaccid areflexia in 60–70 sec • Loss of EEG waves in 10 sec to 2 min, depending on cephalic region (“highest” brain areas failing first)	–
(6)	Dog (n = 31)	To study the recovery of function following arrest of the brain circulation	• Ligature of vertebral arteries and use of a cervical pressure cuff (during 2–11 min, at 350 mm Hg) • Airways open (tracheal intubation)	Loss of respiratory movements in 15–20 sec (never longer than 30 sec)	• Loss of muscle movement (spinal shock) in 1.5–2 min • No convulsions following complete arrest of the brain circulation; in one case of incomplete arrest, convulsions were seen • Urination frequently occurred during the first minute	Strong and rapid pulse during first few minutes, then it weakens
(7)	Dog (n = 16)	To elucidate the nature of the central vasopressor reflex	• Ligature of the internal and external carotids and/or the vertebral arteries • Airways open (tracheal intubation)	• Loss of respiratory movements between 45 sec and 8 min (in 2/3 of animals) • Respiration continued at a slow rate in one-third of animals	Complete medullary paralysis in 8–14 min (spinal animal)	The pulse rate decreased after c. 45 sec

EEG, electroencephalogram.

the brain waves were being recorded. The brain was later cut into gross sections to verify the position and depth of the electrodes.

Kabat et al. (6) used dogs as their animal model in an attempt to achieve complete obstruction of blood flow to the brain while maintaining adequate perfusion to the rest of the body. To avoid the use of anesthesia while observing the reactions of the animals to asphyxia, the dogs underwent a surgical procedure several weeks prior to the experiment to ligate both vertebral arteries. Therefore, at the time of the experiment, no surgery was required. To occlude the carotids, a pressure cuff was inflated around the neck of the dog and the airway remained intact because of the fact that the dogs were orally intubated. The dogs were given a dose of atropine sulfate to prevent vagal cardiac inhibition, and their heart rates were monitored throughout the experiment.

Guyton et al. (7) also used dogs to observe the effects of cerebral ischemia on vasomotor reflexes. The dogs were anaesthetized with sodium pentobarbital. All of the vessels leading to the brain were ligated except the carotids for one series of dogs, and the vertebrals for another set. The carotid sinuses were denervated using a phenol solution, and the lack of electrical activity was confirmed by a lack of response to stimulation. Screw clamps were installed around the patent arteries so that they could be occluded as needed. An intratracheal cannula was used to maintain respiration. Blood pressure was monitored using a femoral cannula.

Animal Studies: Group 2

In 1992, a forensic study implemented a more accurate model using dogs that combined ligature of the trachea with ligature of the vessels of the neck (internal jugular veins, carotid arteries, and vertebral arteries) (Table 2). The methods used by Ikeda attempted to recreate a typical hanging using dogs. Under sodium pentobarbital anesthesia, all of the major structures of the neck were exposed and simultaneous ligation of the vertebral arteries, carotid arteries, internal jugular veins, vagus nerves, and trachea was achieved. During that time, blood pressure, intrathoracic pressure, electrocardiogram, and electroencephalogram were all recorded.

Animal Studies: Group 3

The third group is composed of three animal models, including rabbits or rats hanged by the neck with ropes, to recreate the hanging process (Table 3). In one of these studies, the closure of the upper part of the laryngeal cavity was confirmed after death by analyzing resin castings of the airways (9). In this same study, vertebral arteries remained permeable on postmortem angiographs. The objective of the other two studies in this group was to evaluate the expression of hypoxia-inducible factor-1 alpha (HIF1-alpha), a marker of hypoxia (10,11). As the study was not examining the pathophysiology of hanging, there was no direct confirmation of the degree of occlusion of the neck structures achieved during the experiment.

Sawaguchi developed a restraining and strangling apparatus for rabbits that was designed to simulate a typical hanging because the ligature was tightened by the animal's own body weight (9). Lidocaine solution was used to anesthetize the rabbits. Blood pressure, heart rate, respiratory rate, and electrocardiography were all followed on a cardiopulmonary monitor during the experiment. Cerebral blood flow and brain oxygen tension were measured using a temperature-controlled thermoelectrical tissue blood flow monitor and an oxygen microelectrode, respectively. Following the death of the animal, resin molds of the airway and angiography of the arterial vessels were performed to ensure that complete occlusion had occurred.

Zhang et al. (10,11) used rats to examine the induction of HIF1-alpha in two different models of asphyxia. A third of their rats died of nitrogen gas inhalation, a third died from being hung from a fixed bar with nylon rope, and the control third died from breaking the cervical spine by grabbing the head and tail. Tissue samples were collected from various organs to evaluate the expression of HIF1-alpha using immunohistochemistry and half-quantitative RT-PCR.

Respiratory Responses to Hanging

In animal models of asphyxia with isolated occlusion of the neck vessels (group 1), the loss of respiratory movements was observed in <2 min in all of the studies but one (Fig. 1). In comparison, the stage of apnea was reached after more than 2 min in animal models combining occlusion of neck vessels with airway obstruction (group 2) or models that reproduced hanging by the neck (group 3). Thus, the occlusion of the airways during hanging does not seem to accelerate the decline of respiratory movements in animal models. It should be emphasized that it is difficult to compare the effects of hanging on respiratory movements that were reported in the various studies because of the diversity of animal models that were used, especially with respect to species. However, three different dog models have been developed, allowing for better

TABLE 2—*The course of respiratory, circulatory, and neurological responses following animal asphyxia simulating hanging with obstruction of the vasculature of the neck and airway passages.*

Reference	Model	Objective	Methods	Respiratory Component	Neurological Component	Circulatory Component
(8)	Dog (n = 15)	To compare the course of respiration and circulation in death due to typical hanging versus obstructive asphyxia	• Ligature of bilateral common carotid arteries, vertebral arteries, vagus nerves, and internal jugular veins • Ligature of the trachea	• Immediate increase of respiratory movements, maintained for 1–1.5 min • Delayed cessation of respiratory movements (initial apnea) lasting for 0.5–1 min • Terminal respirations: sporadic respiratory movements with a sharp inspiratory component, lasting 2–3 min • Total duration of respiratory movements: between 4 and 6 min	• Loss of EEG waves in 1.5–2 min • Convulsive waves on EEG during the terminal respirations	• The heart rate increased after 2 or 3 respiratory movements until circulatory collapse occurred at 4–6 min • Increased blood pressure (×1.5) maintained for 1–1.5 min until apnea occurs, then the pressure gradual decreases

EEG, electroencephalogram.

TABLE 3—*The course of respiratory, circulatory, and neurological responses following animal asphyxia simulating hanging with an external ligature tightened around the neck.*

Reference	Model	Objective	Methods	Respiratory Component	Neurological Component	Circulatory Component
(9)	Rabbit (n = 10)	To study the effects of compression of the carotid arteries and obstruction of the airway on cerebral blood flow (CBF) and brain oxygen tension (PtO_2)	Hanging by the neck with a wire rope of 1 mm diameter with a closed loop	• Respiratory arrest (irreversible apnea) in 6–6.3 min • Brain oxygen tension decreased with time reaching 2/3 of its initial value at the time of irreversible apnea	–	• Cerebral blood flow dropped immediately then rose suddenly and progressively dropped after 3–4 min • Terminal hypotension in 6.5–8.5 min • Vertebral arteries remained permeable at angiography
(10)	Rat (n = 20)	To study the expression of hypoxia-inducible factor 1-alpha (HIF1-alpha) in heart and lung tissue following asphyxia	Hanging by the neck with a nylon rope of 3 mm diameter	–	Loss of muscle movement in 1.8–3.5 min (average = 2.9 min)	Absence of heart beat in 4.5–6 min
(11)	Rat (n = 18)	To study the expression of HIF1-alpha in heart, lung, liver, and kidney tissue following asphyxia	Hanging by the neck with a nylon rope of 3 mm diameter	Loss of respiratory movements in 120–197 sec	• Erection of the tail observed • Urination frequently occurred	Absence of heart beat in 250–340 sec

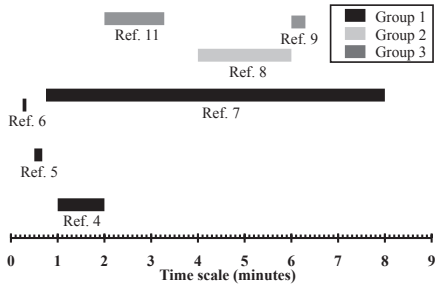


FIG. 1—*Loss of respiratory movements in animal models. Ref., reference; references 4 and 5, Cat models; references 6, 7 and 8, dog models; reference 9, rabbit model; reference 11, rat model.*

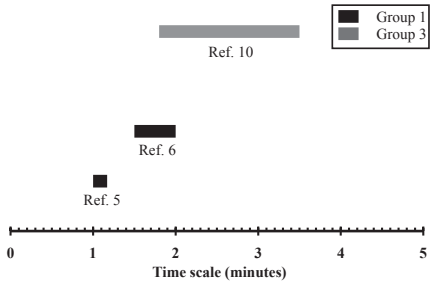


FIG. 2—*Loss of muscle movements in animal models. Ref., reference; reference 5, cat model; reference 6, dog model; reference 10, rat model.*

comparison. In Kabat et al. (6), dogs were asphyxiated by a combination of vertebral artery ligature and external pressure on the neck by a cuff, the airways kept open by tracheal intubation. Airways were also kept open by tracheal intubation in Guyton, but the carotids, vertebrals, and accessory blood vessels were all ligatured (7). In Ikeda et al. (8), a tracheal ligature accompanied the neck vessels ligature. As depicted in Fig. 1, the loss of respiratory movement was not quicker with tracheal occlusion (group 2) compared with isolated neck vessels ligature (group 1).

Deep, rhythmic, abdominal respiratory movements were observed during all of the cases of human hanging analyzed by the Working Group on Human Asphyxia (WGHA) (2,3). These rhythmic respiratory movements generally ended between 1 min 2 sec and 2 min 5 sec. Furthermore, the passage of air through the airways was clearly audible in several filmed hangings, without stridor, confirming that the airways were not completely occluded. Therefore, the loss of respiratory movements does not seem to depend on tracheal obstruction in human hangings. This observation in humans is in keeping with animal models. This similarity between animal and human studies emphasizes that cessation of cerebral blood flow, rather than airway obstruction, is responsible for the loss of respiratory function.

The increased rate of respiration observed in animal models was also reported in an unusual human study by Rossen et al. (12) in 1943. In this study, male volunteers from a prison in Minnesota had special pressure cuffs placed around their necks that compressed the vasculature of the neck, while sparing the tracheal area. Respiration persisted throughout the test, but at an increased rate.

Neurological Responses to Hanging

In animal models of hanging, loss of muscle movement was observed to occur within 1–3.5 min (Fig. 2). This rapid loss of muscle movement was also found in human filmed hangings, the last muscle movement being observed between 1 min 2 sec and 7 min 31 sec (2,3).

Early convulsions were recorded in both of the cat models of asphyxia (4,5). These tonic and clonic generalized convulsions seemed similar to the ones described in human filmed hangings, after 10–19 sec, and in the human experimentation study by Rossen et al. (2,3,12). However, in the dog model of Kabat et al. (6), no convulsions were elicited by the complete arrest of blood circulation to the brain, whereas gradually increasing convulsions culminating to status epilepticus were triggered in a case of incomplete arrest of blood circulation.

The presence of convulsive waves on the electroencephalogram (EEG) was noted in the dog model of Ikeda et al. (8). Other signs of seizure, including tail erection and urination, were mentioned in two animal models (6,11). EEG waves disappeared completely within 2 min in cats and dogs (5,8). In the human study of Rossen et al. (12), EEG showed large, slow waves correlating with the loss

of consciousness. Some patients suffered involuntary urination and defecation.

A complex pattern of decerebrate and decorticate rigidity is depicted in the eight filmed hangings of the WGHA (2,3) and described in a case report of a filmed hanging by Yamasaki et al. (13). No similar findings were reported among the animal studies.

Circulatory Responses to Hanging

Animal models reveal that cardiac function declines during the first few minutes following hanging (4,6,7) and reaches complete circulatory collapse within 4–8.5 min.

It is impossible to compare the circulatory responses induced by asphyxia in animal models to human filmed hangings. However, some comparisons can be drawn from the human study conducted by Rossen (12). Initially, after loss of consciousness, the heart rate was stable. Bradycardia then occurred and the heart rate dropped by 50%. Electrocardiogram changes were minimal.

Avenues to Improve Animal Models of Hanging

The animal models of hanging clearly show similarities with human filmed hangings, particularly when examining the neurological and respiratory responses. Thus, the application of animal models to humans seems to be an interesting path to explore the pathophysiology of hanging. Yet, this review of the eight animal studies reveals that further research is required, especially when considering the interspecies nature of this analysis, the variability of methods used, and the lack of clearly specified objectives.

In humans, as well as in rabbits, rats, and mice, both the internal carotids and the vertebral arteries contribute to cerebral blood flow (14). However, in dogs and pigs, the internal carotids are relatively less developed so it is the vertebral arteries that supply the majority of the cerebral circulation. Furthermore, exceptionally well-developed anastomoses from the external carotids furnish the collateral blood supply to the brain in dogs. This problem of extensive collateral circulation in dogs, mentioned in the study by Guyton (7), makes it difficult to estimate the actual degree of cerebral ischemia induced by carotid and vertebral ligature in his experiments. In cats, the internal carotids are atrophied at birth, and the arterial cerebral circle consists of important epidural vessels (14). Because of these differences in vascular anatomy, rabbits, rats, and mice are the best animals to use to create models of hanging. Considering the size of these animals, the rat may constitute the best model to investigate the pathophysiology of hanging.

Animal models can be broadly divided into three separate groups: isolated ligature of the vasculature of the neck (group 1), combined blood vessel and airway obstruction (group 2), and true hanging of the animal by the neck (group 3). Most of the forensic animal models of hanging that have been developed so far belong to groups 1 and 2. Unfortunately, the validity of these models has yet to be determined because they assume that either vascular occlusion or combined vascular and airway obstruction are the contributors to the mechanism of death. The relative contribution of each theory to the mechanism of death in hanging (occlusion of the airways, occlusion of the neck vessels, and vagal stimulation) has yet to be determined. Therefore, it is recommended to use animal models that reproduce the actual process of hanging itself (group 3), with compression of the neck structures by a rope tightened by the animal's own body weight. Pathophysiological data from this true-hanging model could be later compared with other models reproducing hanging, such as isolated ligature of the blood vessels (group 1) or combined blood vessel and airway obstruction (group 2).

Models of rats hanged by the neck with a rope have been developed so it is already known that the relatively small weight of the rat is sufficient to reproduce the hanging process (10,11). However, these models were developed for molecular biology experiments on H1F1-alpha and were never used in the study of the pathophysiology of hanging.

After hanging the anesthetized animal, the extent to which the vessels are occluded should be evaluated by performing a cerebral angiography and a cervical Doppler. This radiographic test assesses the level of occlusion of the carotids, the vertebral arteries, and the jugular veins (9). Additionally, the use of plethysmography would help to determine the degree of airways obstruction (15–17). This system evaluates respiration by providing flow-volume curves and respiratory rates. Also, a cardiopulmonary monitor would be useful to measure physiological parameters, including ECG, blood pressure, heart rate, and oxygen saturation (9). An EEG can be performed to study changes in cerebral electrical activity and to detect signs of epilepsy (8). Filming the hanging process is strongly encouraged to reexamine the different stages of asphyxia. For obvious ethical reasons, animals have to be anesthetized. Preferably, the drug used should have no proconvulsive or anticonvulsive effect and no respiratory interference. In three previous animal studies, barbiturate drugs were used and may have interfered with the neurologic and respiratory responses (5,7,8). Considering that barbiturate drugs induce respiratory depression and have anticonvulsive effects, they are not recommended for animal models of asphyxia. Ketamine, a short-acting general anesthetic with hallucinogenic and painkilling qualities, could be a more appropriate choice (18,19).

References

1. DiMaio VJ, DiMaio D. Asphyxia In: Forensic pathology. 2nd edn. Boca Raton, FL: CRC Press, 2001;299–377.
2. Sauvageau A. Agonal sequences in four filmed hangings: analysis of respiratory and movement responses to asphyxia by hanging. J Forensic Sci 2009;54(1):192–4.
3. Sauvageau A, LaHarpe R, Geberth VJ. Agonal sequences in eight filmed hangings: analysis of respiratory and movement responses to asphyxia by hanging. J Forensic Sci. E-pub ahead of print. DOI: 10.1111/j.1556-4029.2010.01434.x.
4. Gildea EF, Cobb S. The effects of anemia on the cerebral cortex of the cat. Arch Neurol Psychiatry 1930;23:876–903.
5. Sugar O, Gerard RW. Anoxia and brain potentials. J Neurophysiol 1938;1:53–62.
6. Kabat H, Dennis C, Baker AB. Recovery of function following arrest of the brain circulation. Am J Physiol 1941;182:737–47.
7. Guyton AC. Acute hypertension in dogs with cerebral ischemia. Am J Physiol 1948;154(1):45–54.
8. Ikeda N, Harada A, Suzuki T. The course of respiration and circulation in death due to typical hanging. Int J Legal Med 1992;104(6): 313–5.
9. Sawaguchi A. Medicolegal diagnosis of asphyxia – cerebral blood flow and brain oxygen tension. Nihon Hoigaku Zasshi 1992;46(6):375–8.
10. Zhang GQ, Zhou B, Du B, Yang ZH, Zhang BL, Zhu YH, et al. [Expression of HIF1-alpha on myocardium and lung in rats model of asphyxia death]. [In Chinese] Fa Yi Xue Za Zhi 2006;22(6):407–10.
11. Zhang BL, Yang ZH, Ran P, Liang WB, Zhou B, Zhang GQ, et al. [Induction of hypoxia-inducible factor-1alpha in two kinds of rats asphyxiation death models]. [In Chinese] Fa Yi Xue Za Zhi 2007;1:4–7.
12. Rossen R, Kabat H, Anderson JP. Acute arrest of cerebral circulation in man. Arch Neurol Psychiatry 1943;50:510–28.
13. Yamasaki S, Kobayashi AK, Nishi K. Evaluation of suicide by hanging from the video recording. Forensic Sci Med Pathol 2007;3(1):45–51.
14. Roger T. Anatomie comparée des animaux de laboratoire [Vet-Lyon.fr Web site]. November, 2008, http://www2.vet-lyon.fr/ens/expa/cours/anat-comparee/anatcomp_cardiovasc.htm (accessed March 30, 2009).
15. Glaab T, Taube C, Braun A, Mitzner W. Invasive and noninvasive methods for studying pulmonary function in mice. Respir Res 2007; 8:63.

16. Hoymann HG. Invasive and noninvasive lung function measurements in rodents. J Pharmacol Toxicol Methods 2007;55(1):16–26.
17. O'Neil JJ, Raub JA. Pulmonary function testing in small laboratory mammals. Environ Health Perspect 1984;56:11–22.
18. White PF, Way WL, Trevor AJ. Ketamine – its pharmacology and therapeutic uses. Anesthesiology 1982;56(2):119–36.
19. Bergman SA. Ketamine: review of its pharmacology and its use in pediatric anesthesia. Anesth Prog 1999;46(1):10–20.

Additional information and reprint requests:
Anny Sauvageau, M.D., M.Sc.
Office of the Chief Medical Examiner
7007, 116 Street
Edmonton, Alberta T6H 5R8
Canada
E-mail: anny.sauvageau@gmail.com

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01434.x
Available online at: interscience.wiley.com

PAPER

PATHOLOGY AND BIOLOGY

Anny Sauvageau,[1] M.D., M.Sc.; Romano LaHarpe,[2] M.D.; Vernon J. Geberth,[3] M.S., M.P.S.; and The Working Group on Human Asphyxia

Agonal Sequences in Eight Filmed Hangings: Analysis of Respiratory and Movement Responses to Asphyxia by Hanging*

ABSTRACT: It has been proposed that filmed hangings may hold the key to a better understanding of human asphyxia, and The Working Group on Human Asphyxia was formed to systematically review and compare these video recordings. This study analyzed eight filmed hangings. Considering time 0 to represent the onset of the final hanging, rapid loss of consciousness was observed (at 8–18 sec), closely followed by convulsions (at 10–19 sec). A complex pattern of decerebrate rigidity and decorticate rigidity then followed. Between 1 min 38 sec and 2 min 15 sec, muscle tone seemed to be lost, the body becoming progressively flaccid. From then on, isolated body movements were observed from time to time, the last one occurring between 1 min 2 sec and 7 min 31 sec. As for the respiratory responses, all cases presented deep rhythmic abdominal respiratory movements (last one between 1 min 2 sec and 2 min 5 sec).

KEYWORDS: forensic science, hanging, asphyxia, video recording, pathophysiology, human

Hanging is a form of asphyxia secondary to compression or constriction of the neck structures by a noose or other constricting band tightened by the weight of the body (1). Hanging can be either incomplete (partial suspension) or complete (full suspension), depending on whether or not parts of the body touch the ground (e.g., toes, feet, knees, or buttocks) (1). Death is caused by closure of the blood vessels and/or air passages of the neck, with insufficient oxygen reaching the brain (1). A reflex vagal inhibition by stimulation of the baroreceptors in the carotid sinuses and the carotid body may also play a role. Although theses generalities are found in all forensic textbooks (1–3), further description of the pathophysiology of hanging is very limited. This situation is not surprising considering the paucity of available research: apart from a few animal studies (4), most of our contemporary body of knowledge is in fact based on old writings from the end of the 19th century and beginning of the 20th (5). Judicial hangings (executions) can still be witnessed nowadays in a few countries, but those deaths are very different in nature from typical hangings, death being caused by fracture dislocation of the upper cervical vertebrae with transection of the cord rather than asphyxia by compression of neck structures.

[1]Office of the Chief Medical Examiner, 7007 116 Street, Edmonton, Alberta T6H 5R8, Canada.
[2]Centre Universitaire Romand de Médecine Légale—Site Genève, 1 Michel Servet Street, (CMU), 1211 Geneva 4, Switzerland.
[3]P.H.I. Investigative Consultant Inc, PO Box 197, Garnerville, NY 10923.
*Presented at the 61st Annual Meeting of the American Academy of Forensic Sciences, February 16–21, 2009, in Denver, CO.

Received 31 Mar. 2009; and in revised form 3 Aug. 2009; accepted 8 Aug. 2009.

It is known, however, that some hanging victims film their hangings, mainly in an autoerotic context. It has been proposed that these filmed hangings may hold the key to a better understanding of human asphyxia. The Working Group on Human Asphyxia (WGHA) was formed to systematically review and compare these video recordings (6,7). Each scientist who has such a video or who has access to such a video is welcome to join this group and the video will be added to the ongoing study.

This study by the WGHA described the agonal sequences observed in eight filmed hangings.

Material and Methods

A total of eight filmed hangings were analyzed: six autoerotic accidents and two suicides. All victims were adult white men.

In the first recording, a man recorded his suicide with a video camera. He tied his neck with a padded rope fixed on the rail system of an electric garage door and used the remote control to close the door, therefore hanging himself. His feet were fixed in ski boots, tied with chains to a metal platform. In the second recording, a man masked with woman's underwear hanged himself in his garage, using a traditional hangman's noose made out of thick rope. A large white sheet was spread on the back wall. He hanged himself from a standing position, his knees slightly bent, and his feet touching the floor. In the third recording, a man dressed in a cowboy costume hanged himself in a basement trap by letting go of the nearby ladder (free hanging with a ligature that seems to be a rope). The fourth recording was of a suicide in custody, filmed by a surveillance camera. The victim was kneeling on the ground, a cloth band tied to the cell bars adjacent to the decedent. These first four videos were previously reported (6,7). The fifth video was

filmed in an autoerotic context: the nude man hanged himself almost completely lying down in a prone position, with a rope. On the sixth recording, a man hanged himself with a cloth band, in a standing position with feet on the ground. His face is covered by a hood. In this video, the victim hangs himself on and off several times, applying intermittent neck compression for several minutes before the final fatal hanging. The seventh video shows the autoerotic accidental hanging of a man in a standing position, feet on the ground, with a rope. As for the eighth recording, an autoerotic accident as well, it illustrated the hanging of a nude man with an electric cord, in a standing position in the hallway, feet on the ground.

For each video, we evaluated the time frame of body responses: loss of consciousness, convulsions, decorticate rigidity, decerebrate rigidity, loss of muscle tone, last muscle movement, and respiratory responses. All recording were evaluated jointly by at least two judges, one of the judges having seen all the videos.

Results

The agonal sequences observed in the eight filmed hangings are presented in Table 1. Considering time 0 to represent the onset of the final hanging, rapid loss of consciousness was observed (at 8–18 sec). Loss of consciousness was largely assessed by a close examination of the victim's face, voluntary movements, and body tonus. In two cases (cases 2 and 4), loss of consciousness was not possible to evaluate: in case 2, the victim's face was masked with underwear and in case 4, image quality from the surveillance camera was not optimal enough to estimate this issue adequately.

Loss of consciousness was closely followed by mild convulsions in all cases (at 10–19 sec). Convulsions were generalized, of the tonic-clonic type.

A complex pattern of decerebrate rigidity and decorticate rigidity then followed: decerebrate rigidity is characterized by a full extension of both upper and lower limbs, whereas decorticate rigidity is associated with flexion of the upper limbs combined with extension of lower limbs and trunk. In all cases but one (case 5), the decerebrate rigidity was first noted (at 11–31 sec), followed in most cases by two separates phases of decorticate rigidity. The first phase of decorticate rigidity was relatively sudden and quick (at around 21 sec–1 min 8 sec). In the second phase, the decorticate rigidity developed slowly and was more sustained (at around 34 sec–1 min 32 sec). It should be mentioned that the time of decerebrate rigidity in case 1 presented here is different from that previously reported (6,7): as the judges gain experience in analyzing the agonal sequences of hanging, it was noted that the true decerebrate rigidity in case 1 was missed in the earlier viewing because it was off field.

Between 1 min 38 sec and 2 min 15 sec, muscle tone seemed to be lost, the body becoming progressively flaccid. From then on, isolated body movements were observed from time to time, the last one occurring between 1 min 2 sec and 7 min 31 sec. In cases 7 and 8, we did not have the original recording, and unfortunately the recording was cut too early to allow evaluation of these late responses.

As for the respiratory responses, all cases presented deep rhythmic abdominal respiratory movements. These respiratory movements started between 13 and 24 sec and stopped between 1 min 2 sec and 2 min 5 sec. It is worth emphasizing that these respiratory movements were not only seen but also heard, confirming the passage of air in the airways despite the hanging process.

Discussion

Loss of Consciousness and Convulsions

The rapid loss of consciousness (in 8–18 sec) observed in this study supports the general affirmation found in forensic textbooks that hanging causes unconsciousness in an average of 10 sec (2,3). It is also in keeping with an old study by Rossen et al. (8): inflation of a pressure cuff on the neck of 85 male volunteers caused loss of consciousness in 5–11 sec. Two other filmed hangings are reported in the literature, one also presenting an early loss of consciousness at 10 sec (9), whereas in the other the loss of consciousness is delayed at 55 sec (10).

Loss of consciousness was closely followed by generalized tonic-clonic convulsions (10–19 sec). This correlates with the same type of convulsions observed in the study on male volunteers (8) and in one of the previous filmed hanging (10). However, the other filmed hanging reported fine twitches but no clonic and tonic spasms (9).

Decerebrate Rigidity and Decorticate Rigidity

Decerebrate rigidity indicates lesions of the brainstem caudal to the red nucleus and rostral to the vestibular nuclei (11). The upper and lower limbs are fully extended, with an extension of hips and knees, plantar flexion of feet and toes, internal rotation of the shoulders, extension of the elbows, hyperpronation of the distal parts of the upper limbs with finger extension at the metacarpophalangeal joints and flexion at the interphalangeal joints (11). Decorticate rigidity on the other hand points toward a cerebral cortex

TABLE 1—*Agonal sequences in eight filmed hangings.*

	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8
Movement responses								
Loss of consciousness	13 sec	nd	18 sec	nd	10 sec	8 sec	10 sec	12 sec
Convulsions	15 sec	14 sec	19 sec	18 sec	13 sec	11 sec	10 sec	14 sec
Decerebrate rigidity	19 sec	19 sec	21 sec	nd	1 min 19 sec	31 sec	11 sec	20 sec
Decorticate rigidity (stage 1)	21 sec	1 min 8 sec	1 min	nd	59 sec	33 sec	26 sec	31 sec
Decorticate rigidity (stage 2)	1 min 11 sec	1 min 32 sec	1 min 4 sec	nd	–	–	34 sec	–
Loss of muscle tone	1 min 38 sec	2 min 15 sec	2 min 4 sec	nd	1 min 52 sec	–	nd	nd
Last muscle movement	4 min 10 sec	2 min 47 sec	3 min 1 sec	nd	7 min 31 sec	1 min 2 sec	nd	nd
Respiratory responses—very deep respiratory attempts								
Start	20 sec	21 sec	22 sec	24 sec	13 sec	19 sec	13 sec	16 sec
End	2 min	2 min 3 sec	2 min 4 sec	nd	2 min 5 sec	1 min 2 sec	nd	nd

nd, no data; –, not observed.

impairment and is characterized by flexion of the upper limb combined with extension of the lower limb (11). Despite the term rigidity, decorticate rigidity is more closely related to spasticity: lesions of the premotor areas are associated with an increased muscle tone, more obvious in the extensor muscles of the legs and flexors of the arms.

In all but one case, decerebrate rigidity was observed first, followed by decorticate rigidity. At the present time, there is neither obvious explanation nor hypothesis to explain this fact.

Jugular veins and carotid arteries are more prone to occlusion by neck compression than the more deeply located vertebral arteries. Considering that the brainstem is vascularized by tributaries of the vertebral arteries whereas the premotor areas are vascularized by tributaries from the carotid, it could have been assumed that decerebrate rigidity would appear first. However, this is not the case.

Time Delay of Agonal Responses in Incomplete and Complete Hangings

It is generally thought in the forensic community that the time delay to agonal responses will vary depending on the type of hanging, hangings with complete suspension of the body leading to death more quickly than hangings in which the body is partially supported (e.g., feet or knees on the floor). This indeed would seem logical. However, this study does not support this assumption. As a matter of fact, agonal responses in complete hanging (case 3) appeared no sooner than in incomplete hangings with feet or knees on the ground (cases 1–2, 4, 6–8) or even incomplete hangings lying down (case 5).

The fastest agonal responses were observed in case 6. Interestingly, the decedent in this case intermittently applied the neck compression longer than the other decedents. This may have resulted in an ultimately more rapid hypoxia leading to earlier observed physiologic distress with the final compression.

Respiratory Responses

All cases demonstrated deep rhythmic abdominal respiratory movements. These respiratory movements were not only visualized but were also clearly audible. This fact strongly supports the notion that vascular occlusion is the major component of the pathophysiology of hanging. It should be pointed out that this last assertion is controversial and further studies are needed before concluding this old debate on the relative contribution of the three main possible mechanisms to death by hanging (occlusion of the airways, occlusion of the neck vessels, and vagal inhibition). It may be argued that the diaphragm and chest wall appear to move as if the person is breathing, without air actually entering or exiting the lungs. The fact that abdominal "breathing" movements are visualized does not mean that air exchange is occurring. However, the breathing movements were not only seen but also heard. It could also be argued that hearing breath sounds does not eliminate the possibility of significant airway obstruction. For example, many choking deaths with airway obstruction by food or foreign objects present with partial pathway for air to move around the obstruction, yet these people still presumably die from airway obstruction. Similarly, if the trachea is compressed by a noose around the neck, there may be substantial narrowing of the lumen but some air could theoretically continue to pass by the area of compression, thus accounting for some degree of actual air passage. Therefore, the mere fact that you can hear breath sounds does not definitely rule out respiratory obstruction as a mechanism of death in these cases. Ultimately, this study strongly supports the idea that tracheal occlusion is not complete in some types of hanging, but it would be premature to totally exclude some implication of partial airways obstruction in the mechanism of death by hanging.

Nevertheless, the presence of clear audible respiration confirming the absence of total airways obstruction in several of these hangings is particularly interesting. For example, case 1 will be further analyzed. In this filmed hanging, the victim is an adult man of 148 pounds (67.3 kg). The hanging is incomplete, with the victim standing, feet on the ground. The amount of body weight involved tightening the ligature in relation to the hanging position can be evaluated by using results from a previous study by Khokhlov (12): more than 65% for a hanging in a standing position feet flat. In case 1, 65% of body weight is about 96 pounds (44 kg). This pressure around the neck is highly superior to the weight of 33 pounds (15 kg) mentioned in textbooks as the pressure necessary to occlude the trachea (1,13). Nevertheless, filmed hanging #1 was clearly associated with audible respiration. The explanation is probably that one important factor is often overlooked: the angle of the pressure on the structures of the neck. As a matter of fact, the weight necessary to occlude the various structures of the neck seems to have been studied with pressure vectors applied relatively perpendicularly to these structures compare to a true typical hanging. Pressure vectors in real life are more diagonal, with a greater angle to the neck structures. Furthermore, another factor may have been underestimated: the level of the ligature on the neck (across the trachea, the larynx, or above the larynx).

Conclusion

Despite differences in the types of hanging in these eight films, similarities in the agonal responses to hanging are striking. Filmed hangings seem to be one key to a better understanding of human asphyxia, and the WGHA will continue to add new cases in the following years.

References

1. DiMaio VJ, DiMaio D. Asphyxia. Forensic pathology, 2nd edn. Boca Raton, FL: CRC Press, 2001;229–77.
2. Spitz WU. Asphyxia. In: Spitz WU, Spitz DJ, editors. Spitz and Fisher's medicolegal investigation of death: guidelines for the application of pathology to crime investigation, 4th edn. Springfield, IL: Charles C Thomas Publisher LTD, 2006;783–845.
3. Saukko P, Knight B. Suffocation and 'asphyxia'. In: Ueberberg A, project editor. Knight's forensic pathology, 3rd edn. London: Arnold Publishers, 2004;352–67.
4. Boghossian E, Clement R, Sauvageau A. Respiratory, circulatory and neurological responses to asphyxia by hanging: a review of animal models. J Forensic Sci. E-pub ahead of print. DOI: 10.1111/j.1556-4029.2010.01436.x.
5. Clement R, Redpath M, Sauvageau A. Cause of death in hanging: a historical review of the evolution of pathophysiological hypotheses. J Forensic Sci. E-pub ahead of print. DOI: 10.1111/j.1556-4029.2010.01435.x.
6. Sauvageau A, Racette S. Agonal sequences in a filmed suicidal hangings: analysis of respiratory and movement responses to asphyxia by hanging. J Forensic Sci 2007;52(4):957–9.
7. Sauvageau A. Agonal sequences in four filmed hangings: analysis of respiratory and movement responses to asphyxia by hanging. J Forensic Sci 2009;54(1):192–4.
8. Rossen R, Kabat H, Anderson JP. Acute arrest of cerebral circulation in man. Arch Neurol Psychiatry 1943;50:510–28.
9. Yamasaki S, Kobayashi AK, Nishi K. Evaluation of suicide by hanging from the video recording. Forensic Sci Med Pathol 2007;3(1):45–51.

10. Risse M, Weiler G. Agonale und supravitale bewegungsabläufe beim erhängen. Beitr Gerichtl Med 1989;47:243–6.
11. Ghez C. Posture. In: Kandel E, Schwartz JH, Jessel TM, editors. Principles of neural science, 3rd edn. Norwalk, CT: Appleton & Lange, 1991;596–607.
12. Khokhlov VD. Calculation of tension exerted on a ligature in incomplete hanging. Forensic Med Pathol 2001;123:172–7.
13. Shkrum M, Ramsay DA. Asphyxia. In: Karch SB, series editor. Forensic pathology of trauma: common problems for the pathologist, 1st edn. Totowa, NJ: Humana Press, 2007;65–179.

Additional information and reprint requests:
Anny Sauvageau, M.D., M.Sc.
Office of the Chief Medical Examiner
7007, 116 Street
Edmonton (Alberta) T6H 5R8
Canada
E-mail: anny.sauvageau@gmail.com

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01448.x
Available online at: interscience.wiley.com

PAPER

PSYCHIATRY & BEHAVIORAL SCIENCES; GENERAL

Heng Choon (Oliver) Chan,[1,2] *M.A.; Wade C. Myers,*[3] *M.D.; and Kathleen M. Heide,*[1] *Ph.D.*

An Empirical Analysis of 30 Years of U.S. Juvenile and Adult Sexual Homicide Offender Data: Race and Age Differences in the Victim–Offender Relationship

ABSTRACT: Little is known about the racial patterns of crimes committed by sexual homicide offenders (SHOs). This study examined race and age influences on victim–offender relationship for juvenile and adult SHOs. A large sample (N = 3868) from the Supplemental Homicide Reports (1976–2005) was used. Analyses of victim–offender patterns included examining victim age effects (child, adolescent, adult, and elderly). The findings revealed several race- and age-based differences. Black offenders were significantly overrepresented in the SHO population. This finding held for juveniles and adults independently. White SHOs were highly likely to kill within their race, "intra-racially" (range 91–100%) across four victim age categories, whereas Black SHOs killed both intra-racially (range 24–82%) and inter-racially (18–76%), with the likelihood of their killing inter-racially increasing as the age of the victim increased. This study underscores the importance of considering victim–offender racial patterns in sexual murder investigations, and it offers practical implications for offender profiling.

KEYWORDS : forensic science, sexual homicide, homicide, crime victims, rape, race, offender profiling

In a systematic public survey of crime seriousness in North America, sexual homicide was ranked second only to the bombing of a public building that killed 20 people (1). Although sexual homicides have attracted extensive media attention over the years, this type of violent crime is rare. Only 0.8% of over 470,000 known U.S. homicides from 1976 to 2004 appeared to be sexual homicides (2). Furthermore, this percentage was found to be relatively stable over the years, and is consistent with the 0.5% rate found over a 10-year period in Virginia (3), and the 0.6% rate found over a 10-year period in Florida for juvenile sexual homicides (4). Estimates of the proportion of U.S. homicides that were sexual homicides for the 5-year period from 1991 through 1995 similarly ranged from 0.6 to 0.9 (5).

Because of the relative rarity of this form of crime along with only a small number of researchers exploring it, sexual homicide remains a fertile area for future study. In the last 20 years, there have been <40 empirically published studies on sexual homicide (6). Accordingly, many areas in the study of sexual homicide are still largely unexplored, including racial differences. Contributing to the understudied nature of this subject is the absence of a standardized definition for sexual homicide, and in turn the unavailability of dependable governmental statistics. In the Uniform Crime Reports (UCRs), the official U.S. national crime statistics source, sexual homicide is indexed under the "Unknown Motive" category. This practice occurred because sexual homicide is typically treated as an "ordinary" homicide and not as a sex crime by most of the law enforcement agencies in North America and the United Kingdom (7–10). Complicating this challenge, the sexual dynamics of such crimes are not always apparent at crime scenes, leading to an underreporting of sexual murders (8,11).

A recent empirical study on sexual homicide using a large sample of U.S. data over a span of 29 years was conducted by recoding data from the UCR Supplemental Homicide Reports (SHRs). Homicides that indicated rape-murder and other sex offense-murder were defined as sexual homicides (2). The identification of sexual homicide cases using this method may include some nonsexual homicide cases that were miscoded and exclude others that were sexual homicides but not coded correctly. Additionally, it is dependent on the homicides actually being identified as such and reported by law enforcement to the Federal Bureau of Investigation (FBI), and a small percentage of cases go underreported in this voluntary process. Despite these imperfections, we nonetheless believe this data source to be the best one currently available for a longitudinal study on sexual homicide in the United States given its large sample.

To date, very few studies have investigated the differences between juvenile and adult sexual murderers in terms of the offending process, offender's characteristics, and victimology. Hill, Habermann, Klusmann, Berner, and Briken (12) were among the first to conduct studies that involved juvenile and adult sexual murderers. They recently published the first follow-up study on sexual murderers (N = 166) for a 10-year period after their release

[1]Department of Criminology, University of South Florida, 4202 E. Fowler Avenue SOC 107, Tampa, FL 33620-8100.
[2]Department of Applied Social Studies, City University of Hong Kong, Kowloon Tong, Kowloon, Hong Kong, S.A.R.
[3]Department of Psychiatry, Brown University, 593 Eddy Street, APC 978, Providence, RI 02903.

Received 5 May 2009; and in revised form 12 July 2009; accepted 23 July 2009.

from incarceration. Eleven percent of the sample subjects ($N = 18$) were juveniles. They found that the recidivism rate of the sexual homicide perpetrators after 20 years at risk was 23.1% for sexual and 18.3% for nonsexual violent offenses. Sexual homicides committed by those under 18 resulted in higher sexual recidivism rates.

Using recoded SHR data, ($N = 3845$), Chan and Heide (2) examined these two distinctive offender age groups in terms of murder weapon selection from 1976 to 2004. Their findings indicated that both juvenile ($N = 452$) and adult sexual murderers ($N = 3393$) were equally likely to use personal weapons and contact and edged weapons in their sex killings. Juveniles, however, were more likely than their adult counterparts to use firearms as their murder weapons in sexual homicides. Juvenile and adult sex killers showed a preference for using personal weapons when they murdered vulnerable victims like children and elderly victims. Contact and edged weapons were more likely to be used against adolescent and adult victims by adult sexual murderers. Juvenile sex killers, however, were more likely to select contact and edged weapons as their murder weapons in killing adolescent victims. Interestingly, firearms were preferred by the juveniles when the sexual homicide victims were adults, consistent with Heide's physical strength hypothesis (13,14).

Literature Review

To provide for a more expansive background to the study at hand, pertinent literature on racial and age differences in homicide, sex crimes, and sexual homicide will be reviewed in the following three sections.

Racial Differences in Nonsexual Homicide

Despite recent declines in homicide, Black offending and victimization rates for homicide have remained much higher than those for their White counterparts, given the low representation of Blacks in the overall U.S. population (15–23). Blacks comprised about 50% of all homicide arrestees, whereas this racial group made up *c.* 13% of the total population in the United States (17,24).

Homicide is an unequivocally intra-racial violent offense; victims and offenders from the same race make up 86–90% of all known homicides (15,25,26). Like single-victim homicides, serial murders are also more likely to be intra-racial killings (27). The percentages of Black homicides that were intra-racial killings were higher than their White counterparts (95–97% vs. 87–91%, respectively) (23,28). According to Phillips (29), the high incidence of intra-racial killings was because of social isolation and residential segregation that limit the social contact among groups of different class and racial backgrounds.

Close to 60% of Black offender–Black victim homicides in Humphrey and Palmer's (26) study involved a primary relationship between the offender and victim when compared to 48% in White intra-racial killings. Black victims who were murdered by their friends were more likely to come from their own racial group. In contrast, White victims were two times more likely to be killed by strangers than Black victims (26).

Inter-racial homicides, killings that cross racial lines, are likely to occur more often in larger cities with higher total homicide rates than in smaller cities, and rural and suburban areas (30). The rate of inter-racial killings depends on the social contacts between the racial groups. If the percentage of a particular racial group in a city is low, then it will increase the probability of inter-racial group interactions, which leads to the higher likelihood of inter-racial killings (30). In a study by Humphrey and Palmer (26), 14% of all homicides were inter-racial homicides, most commonly Black offenders killing White victims (11%), with the converse of White offenders killing Black victims being only 2%.

According to Jacobs and Wood (30), most of the inter-racial homicides of White offenders–Black victims occurred during interpersonal disputes, whereas murders of Black offenders–White victims largely happened during felonies. Wilbanks (23) argued that Black offenders were far more likely than White offenders to murder over monetary reasons like gambling and to kill in gang-related incidents. Black offender inter-racial homicides were believed to be the result of the deprivation of resources and competition for jobs and political power that fueled the racial antagonism against Whites (31). Most of the victims of inter-racial homicides were strangers or acquaintances to their offenders (26).

Racial Differences in Sex Crime

No detailed national statistics on sexual victimization were collected before the year 1973 (32). Currently, the National Crime Victimization Survey (NCVS), as a replacement of the National Crime Survey in 1992, and the UCRs are the two primary official annual national statistics for sexual victimization in the United States. These annual figures on sexual assaults are inordinately low compared to actual incidence because most victims of sexual assault do not report their victimization to the law enforcement officials (33). Relying on available figures, the United States has the highest reported sexual assault cases in comparison with other developed countries like Canada, England, West Germany, Sweden, Denmark, France, Holland, Belgium, and Japan (34).

Rape is most often an intra-racial violent crime (32,35). Results of a study conducted in San Francisco suggested that *c.* 24% of women will be sexually assaulted and 20% will experience an attempted rape in their lifetime. The majority of sexual assault victims are adolescents 12–17 years of age (36). African-American females may be at even higher risk for rape than females from other racial groups, especially marital rape (37,38). Although it is often believed that the majority of rapists are Blacks (39), research has revealed that six in every 10 sex offenders are White males who are in their early twenties (36). Young Black sex offenders, however, are more often arrested for forcible/stranger rapes (35). This finding is consistent with data showing that, for male against female sexual assaults and rapes combined, Black offenders are more likely than White offenders to assault strangers (44% vs. 28% of victimizations) (33,40).

Racial Differences in Sexual Homicide

The available data in this area, hampered in most cases by small, nonrepresentative samples, are limited. Most sexual homicidal offenders in these studies were White (60–95%) and committed intra-racial killings (41–45). Same race victim selection was particularly true in a study of 14 juvenile sexual murderers, most of whom were White (46); 86% selected intra-racial victims.

The literature on Black sexual murderers, although sparse, reveals some anomalies. A landmark clinical report of a 22-year-old African-American male who sexually assaulted and killed six female African-America strangers was among the first studies to document sexual killings, as well as serial killings, by Black offenders (47). In contrast to the intra-racial nature of these killings, 82% (48) and 77% (49) of two predominantly adult-aged samples of Black sexual murderers (sample sizes of 33 and 110) killed

elderly females inter-racially: These elderly females were all White (48) or a combination of White and Hispanic (49).

To the authors' knowledge, no study has investigated race and age differences in victim–offender relationship for juvenile and adult sexual murderers. The current study examines these relationships using a large national sample of sexual murderers.

Method

Thirty years (1976–2005) of the FBI's Supplementary Homicide Report data (see 50) were used in this study. This database contains demographic information on the victim and offender characteristics in homicidal events that were reported to the FBI by participating law enforcement agencies across the nation. Sample subjects included in this study were individuals arrested for homicide with apparent sexual elements that occurred as a result of rape and other sexual offenses. Out of a total of 597,351 individuals arrested for homicide over the period of 30 years under review, 3868 (0.6%) of cases with pertinent offense-related information were categorized as sexual homicides.

The age of the sexual homicide offender (hereafter abbreviated as "SHO") was one of the two primary offender variables examined in this study. Juveniles were defined as under age 18; adults as 18 years of age and over. The second variable of analysis was the racial group of the SHO. Race differences were investigated by focusing on the two racial groups that comprised 98.1% of all arrests for sexual homicide in the United States: (i) White and (ii) Black. The two other offender racial groups (Asian and Pacific Islander, American Indian or Alaskan Native) coded by the FBI were excluded in this study because of the limited involvement of these racial groups (1.9%) in sexual homicide arrests. The SHR database does not code offenders or victims as multiracial and no longer records Hispanic origin. Accordingly, analyses by multiracial or Hispanic group affiliations are not possible with the SHR dataset.

Four categories were created for victim type: (i) child, age 12 years and below, (ii) adolescent, age 13–17 years, (iii) adult, age 18–59 years, and (iv) elderly, age 60 years and above. Two racial groups were coded for victim type: (i) White and (ii) Black. Victims from other races constituted only 2.6% of the overall victim population and were also excluded from the current study.

This study served two broad aims. It was designed to explore the characteristics of sexual homicide crimes and SHOs using a large national database that consisted of those apprehended for homicide over the 30-year period from 1976 to 2005. In addition to being exploratory, this study had an analytical component. Chi-square analyses were performed to compare victim–offender racial relationships by different offender age and racial groups across different victim age and racial groups. Significance level was set at 0.05.

There are two ways of looking at racial differences in victim–offender data in sex murders. One can focus on the victims, asking the question: By whom are White and Black victims being murdered? Conversely, one can focus on the offenders, asking who are White and Black offenders killing? From the standpoint of offender profiling, the answers to both questions are important. Accordingly, data from chi-square analyses pertinent to both questions are presented.

Characteristics of Sample Subjects

Analyses of sample demographics were based on the available SHR data with known cases for offender's race ($N = 3868$), offender's age according to race ($N = 3792$), victim's race ($N = 3787$), and victim's age ($N = 3836$). Over a span of three decades, of all those Black and White offenders apprehended for sexual homicide, 59% were White, while the remaining 41% of those arrested were Black. These racial percentages remained the same for juvenile and adult SHO groups independently.

As depicted in Table 1, 88% of all SHOs arrested were over age 18; 52% of the entire sample were White adults; and 36% were Black adults. Juvenile sexual murderers comprised 12% of all sample SHOs, 7% being White juveniles and 5% being Black juveniles.

In terms of victim characteristics, nearly three-quarters (72%) of the victims were White, and the remaining 28% were Black. About three-quarters (77%) of the victims of sexual murderers were over 18 years (64% adult; 13% elderly) and slightly less than one-quarter (23%) were under 18 years (11% children; 12% adolescents).

Results

Chi-square analyses proceeded in eight stages. The first four analyses examined the relationship between offender race and victim race, offender race and victim race controlling for offender age, and the relationship between offender race and victim age group. The next four analyses explored the relationship between offender race and the race of the victim within each of the victim age categories (child, adolescent, adult, and elderly), controlling for the effect of offender age.

Victim–Offender Racial Differences in Sexual Homicides

As shown in Table 2, victim race differed significantly by offender race ($\chi^2(1) = 1459.06$, $p < 0.001$). Although both racial groups were likely to be killed by members of their own race, Black victims were more likely to be murdered by members of their own race (90%) than White victims (78%). Consistent with the victimization data, the pattern of offending shows sharp racial differences. White SHOs killed within their race 95% of the time. In contrast, Black SHOs killed intra-racially only 61% of the time. Additionally, Black SHOs were significantly overrepresented (41%) in the SHO population given their 13% representation in the U.S. population ($\chi^2(1) = 53.02$, $p < 0.001$).

TABLE 1—*Variables of the sexual homicide sample extracted from the Uniform Crime Reports [United States]: Supplementary Homicide Reports, 1976–2005 (*N = *3868).*

	Number of Cases	Percent of Total (100%)
Sexual murderer race	($N = 3868$)	
White	2286	59%
Black	1582	41%
Sexual murderer age by race	($N = 3792$)	
Juvenile White	263	7%
Juvenile Black	182	5%
Adult White	1990	52%
Adult Black	1357	36%
Victim race	($N = 3787$)	
White	2739	72%
Black	1048	28%
Victim age group	($N = 3836$)	
Child	441	11%
Adolescent	473	12%
Adult	2439	64%
Elderly	483	13%

TABLE 2—*Victim–offender racial relationship (*N = *3787).*

	Sexual Murderer Race		
Victim Race	White	Black	Total (%)
White	2131	608	2739
Row percent	78%	22%	100%
Column percent	95%	39%	
Black	100	948	1048
Row percent	10%	90%	100%
Column percent	5%	61%	
Total (%)	2231	1556	3787
Row percent	59%	41%	100%
Column percent	100%	100%	

$\chi^2(1) = 1459.06$, Phi = 0.62, $p < 0.001$.

Victim–Offender Racial Differences in Sexual Homicides by Juvenile and Adult Offenders

When adult and juvenile samples were independently examined, as in Table 3, Black SHOs were again significantly overrepresented among juvenile SHOs ($\chi^2(1) = 24.11$, $p < 0.001$) given their 15% representation in the U.S. population for those under the age of 18 (51). Similarly, Black SHOs were also significantly overrepresented among adult SHOs ($\chi^2(1) = 60.55$, $p < 0.001$) given their 12% representation in the U.S. population for those 18 years and above (52).

Juvenile White and Black SHOs differed significantly according to the race of victim they targeted ($\chi^2(1) = 125.80$, $p < 0.001$). Table 3 shows that both White and Black victims were likely to be killed intra-racially, with Black victims being more likely to be killed by members of their own group (86%) than White victims (75%). Offender racial differences in victim selection are pronounced when the analysis focuses on the race of the victims whom juvenile SHOs killed. Nearly 19 of 20 White juvenile SHOs killed victims within their own race, whereas only 11 of 20 Black juvenile SHOs killed within their own race.

Adult White and Black SHOs also differed significantly with respect to the race of their victims ($\chi^2(1) = 1334.23$, $p < 0.001$).

TABLE 3—*Victim–offender racial relationship of juvenile sexual murderer (*N = *436) and adult sexual murderer race (*N = *3277).*

	Offender Race		
Victim Race	White	Black	Total (%)
Juvenile sexual murderer ($\chi^2(1) = 125.71$, Phi = 0.54, $p < 0.001$)			
White	242	82	324
Row percent	75%	25%	100%
Column percent	94%	46%	
Black	16	96	112
Row percent	14%	86%	100%
Column percent	6%	54%	
Total (%)	263	181	436
Row percent	59%	41%	100%
Column percent	100%	100%	
Adult sexual murderer ($\chi^2(1) = 1334.23$, Phi = 0.64, $p < 0.001$)			
White	1860	502	2362
Row percent	79%	21%	100%
Column percent	95%	38%	
Black	81	834	915
Row percent	9%	91%	100%
Column percent	4%	62%	
Total (%)	1979	1355	3277
Row percent	59%	41%	100%
Column percent	100%	100%	

As depicted in Table 3, a similar killing pattern is observed with respect to adult sexual murderers. Black victims were significantly more likely to be killed intra-racially (91%) than White victims (79%). Similar to their juvenile counterparts, adult Black SHOs were significantly more likely than White adult SHOs to kill outside their race (38% vs. 4%).

Types of Sexual Homicide Victims by Offender Race

As shown in Table 4, the types of victims killed by White and Black SHOs differed significantly ($\chi^2(3) = 55.83$, $p < 0.001$). Adults constituted the largest victim group for both White SHOs (63%) and Black SHOs (64%); differences in these percentages were not significant. Significant differences were found between White and Black SHOs, however, in the killing of children ($\chi^2 = 10.53$, df = 1, $p < 0.05$), adolescents ($\chi^2 = 9.87$, df = 1, $p < 0.05$), and elderly victims ($\chi^2 = 35.41$, df = 1, $p < 0.001$). Children and adolescents were twice as likely to be killed by White than Black SHOs. In contrast, elderly victims were significantly more likely to be killed by Black SHOs than White SHOs (54% vs. 46%).

Differences in victim selection by White and Black SHOs were again notable. When compared to their racial counterparts, White SHOs were significantly more likely to kill children (13% vs. 9%) and adolescents (14% vs. 10%), whereas Black SHOs were significantly more likely to kill elderly victims (17% vs. 10%).

Types of Sexual Homicide Victims by Offender Race Controlling for Offender Age

Table 5 reveals that the types of victims killed by juvenile White and Black SHOs differed significantly ($\chi^2(3) = 18.57$, $p < 0.001$). Adult victims were the most common class of victims chosen by juvenile White (42%) and Black (51%) offenders. Adolescents (28%) were the next frequent victim age group to be killed by juvenile White SHOs, whereas juvenile Black SHOs selected elderly victims (21%) as their second most common victim class. Only the difference between juvenile White and Black SHOs in the killing of adolescents, however, was significant ($\chi^2 = 11.07$, df = 1,

TABLE 4—*Victim age group by sexual murderer race (*N = *3836).*

	Sexual Murderer Race		
Victim Age Group	White	Black	Total (%)
Child*	294	147	441
Row percent	67%	33%	100%
Column percent	13%	9%	
Adolescent†	313	160	473
Row percent	66%	34%	100%
Column percent	14%	10%	
Adult	1438	1001	2439
Row percent	59%	41%	100%
Column percent	63%	64%	
Elderly‡	221	262	483
Row percent	46%	54%	100%
Column percent	10%	17%	
Total (%)	2266	1570	3836
Row percent	59%	41%	100%
Column percent	100%	100%	

$\chi^2(3) = 55.83$, Cramer's $V = 0.12$, $p < 0.001$.
*Child victims ($\chi^2 = 10.53$, df = 1, $p < 0.05$).
†Adolescent victims ($\chi^2 = 9.87$, df = 1, $p < 0.05$).
‡Elderly victims ($\chi^2 = 35.41$, df = 1, $p < 0.001$).

TABLE 5—*Victim age group by juvenile sexual murderer race (*N = *444) and adult sexual murderer race (*N = *3319).*

Victim Age Group	Offender Race: White	Black	Total (%)
Juvenile sexual murderer ($\chi^2(3) = 18.57$, Cramer's $V = 0.20$, $p < 0.001$)			
Child	46	28	74
Row percent	62%	38%	100%
Column percent	18%	15%	
Adolescent*	74	24	98
Row percent	75%	25%	100%
Column percent	28%	13%	
Adult	111	92	203
Row percent	55%	45%	100%
Column percent	42%	51%	
Elderly	31	38	69
Row percent	45%	55%	100%
Column percent	12%	21%	
Total (%)	262	182	444
Row percent	59%	41%	100%
Column percent	100%	100%	
Adult sexual murderer ($\chi^2(3) = 38.81$, Cramer's $V = 0.11$, $p < 0.001$)			
Child†	248	118	366
Row percent	68%	32%	100%
Column percent	13%	9%	
Adolescent	238	134	372
Row percent	64%	36%	100%
Column percent	12%	10%	
Adult	1298	884	2182
Row percent	60%	40%	100%
Column percent	66%	65%	
Elderly‡	188	211	399
Row percent	47%	53%	100%
Column percent	9%	16%	
Total (%)	1972	1347	3319
Row percent	59%	41%	100%
Column percent	100%	100%	

*Adolescent victims ($\chi^2 = 11.07$, df = 1, $p < 0.05$).
†Child victims ($\chi^2 = 10.54$, df = 1, $p < 0.05$).
‡Elderly victims ($\chi^2 = 25.06$, df = 1, $p < 0.001$).

$p < 0.05$). Nearly three-quarters of the 98 adolescents killed in sexual murders were killed by White SHOs.

Like their juvenile counterparts, adult White and Black SHOs also differed significantly in the killing of different types of victims ($\chi^2(3) = 38.81$, $p < 0.001$). Racial differences were significant with respect to children and elderly victims killed. More than two-thirds of children killed in sexual homicides by adult offenders were killed by White offenders ($\chi^2 = 10.54$, df = 1, $p < 0.05$). Although no significant differences were found, adolescents and adults were also more likely to be killed by White SHOs (64% and 60%, respectively) than their Black counterparts (36% and 40%). In contrast, more elderly victims were killed by Black SHOs than White SHOs (53% vs. 47%) ($\chi^2 = 25.06$, df = 1, $p < 0.001$). Similar to the juvenile Black sexual homicide offending pattern, elderly victims (16%) were the next frequently selected victim age group behind the adult victim age group (65%) to be killed by adult Black SHOs. Interestingly, in contrast to their juvenile White sexual homicide offending, children (13%) were the next frequently killed victim age group after the adult victim age group (66%) by adult White sex killers.

Specific Types of Sexual Homicide Victims by Offender Race Within Offender Age Categories

Analyses examined differences between juvenile and adult sexual murderers by offender race within each of the victim types (child, adolescent, adult, and elderly victims). Several racial differences emerged.

TABLE 6—*Child victim race by juvenile sexual murderer race (*N = *74) and adult sexual murderer race (*N = *360).*

Child Victim Race	Offender Race: White	Black	Total (%)
Juvenile sexual murderer ($\chi^2(1) = 43.67$, Phi = 0.77, $p < 0.001$)			
White	43	5	48
Row percent	90%	10%	100%
Column percent	94%	18%	
Black	3	23	26
Row percent	12%	88%	100%
Column percent	6%	82%	
Total (%)	46	28	74
Row percent	62%	38%	100%
Column percent	100%	100%	
Adult sexual murderer ($\chi^2(1) = 223.04$, Phi = 0.79, $p < 0.001$)			
White	232	21	253
Row percent	92%	8%	100%
Column percent	95%	18%	
Black	12	95	107
Row percent	11%	89%	100%
Column percent	5%	82%	
Total (%)	244	116	360
Row percent	68%	32%	100%
Column percent	100%	100%	

Child Victim Race Results

As depicted in Table 6, significant differences were found between juvenile White and Black SHOs in the killing of White and Black children ($\chi^2(1) = 43.67$, $p < 0.001$). White children were significantly more likely to be killed by juvenile SHOs from their own racial group (90%) than were Black children (88%). However, the differences in the percentages were rather small. In contrast to victim data, examination of offender data reveals noticeable racial differences with respect to child victims. Black SHOs relative to White SHOs were three times more likely to target children outside of their race (18% vs. 6%).

Significant racial differences were also found in the killing of children by adult White and Black SHOs ($\chi^2(1) = 223.04$, $p < 0.001$). The patterns observed are very similar to those observed with respect to juvenile SHOs. Although White children were significantly more likely to be murdered by adult SHOs within their race (92%) than their Black counterparts (89%), the difference is rather small. Once again, however, offender data revealed startling findings, very consistent with those found with respect to juvenile SHOs. Adult Black SHOs were 3.5 times more likely to kill victims outside of their race than adult White SHOs (18% vs. 5%).

Adolescent Victim Race Results

When adolescent victims of juvenile sexual murderers are examined, significant racial differences were also found between juvenile White and Black SHOs ($\chi^2(1) = 44.60$, $p < 0.001$). Whites in the age of 13–17 years were significantly more likely to be murdered intra-racially (91%) by juvenile SHOs than Black adolescents (83%). Dramatic differences are discernible when offending data are examined. Juvenile Black SHOs were eight times more likely to sexually murder adolescents outside their race than White SHOs (32% vs. 4%).

Adult White and Black SHOs also differed significantly in the killing of adolescents when victim and offender race were

examined ($\chi^2(1) = 214.58$, $p < 0.001$). Although Black and White adolescent victims were more likely to be killed by SHOs within their own racial group, the pattern observed with respect to adult offenders is different from the one observed with respect to juvenile offenders. White adolescents were noticeably less likely than Black adolescents to be murdered intra-racially by adult SHOs (86% vs. 98%). Offender data once again highlight large racial differences. Adult Black SHOs were far more likely to kill outside of their race than their White counterparts (29% vs. 1%) (Table 7).

Adult Victim Race Results

Table 8 reveals that races of adult sexual homicide victims differed significantly by juvenile White and Black SHOs ($\chi^2(1) = 47.59$, $p < 0.001$). Both victim and offender data reveal striking racial differences. Although both adult White and Black victims were likely to be killed intra-racially by juvenile SHOs, adult White victims were significantly less likely to be killed intra-racially (70%) than their Black counterparts (83%). Examination of offender data reveals that juvenile Black SHOs were about five times more likely to kill adult victims outside their race than their White counterparts (46% vs. 9%).

Similarly, significant differences were found in the killing of adult victims of different races by adult White and Black SHOs ($\chi^2(1) = 849.75$, $p < 0.001$). The same pattern in intra-racial victimization is seen with respect to adult victims, with White victims less likely to be killed by members of their own race (79%) than Black victims (90%). Offender data indicate that adult Black SHOs were more than seven times as likely to murder victims outside their race as adult White SHOs (37% vs. 5%).

TABLE 7—*Adolescent victim race by juvenile sexual murderer race (*N = *94) and adult sexual murderer race (*N = *365).*

	Offender Race		
Adolescent Victim Race	White	Black	Total (%)
Juvenile sexual murderer ($\chi^2(1) = 44.60$, Phi = 0.69, $p < 0.001$)			
White	69	7	76
Row percent	91%	9%	100%
Column percent	96%	32%	
Black	3	15	18
Row percent	17%	83%	100%
Column percent	4%	68%	
Total (%)	72	22	94
Row percent	77%	23%	100%
Column percent	100%	100%	
Adult sexual murderer ($\chi^2(1) = 214.58$, Phi = 0.77, $p < 0.001$)			
White	232	38	270
Row percent	86%	14%	100%
Column percent	99%	29%	
Black	2	93	95
Row percent	2%	98%	100%
Column percent	1%	71%	
Total (%)	234	131	365
Row percent	64%	36%	100%
Column percent	100%	100%	

TABLE 8—*Adult victim race by juvenile sexual murderer race (*N = *200) and adult sexual murderer race (*N = *2131).*

	Offender Race		
Adult Victim Race	White	Black	Total (%)
Juvenile sexual murderer ($\chi^2(1) = 47.59$, Phi = 0.49, $p < 0.001$)			
White	99	42	141
Row percent	70%	30%	100%
Column percent	91%	46%	
Black	10	49	59
Row percent	17%	83%	100%
Column percent	9%	54%	
Total (%)	109	91	200
Row percent	55%	45%	100%
Column percent	100%	100%	
Adult sexual murderer ($\chi^2(1) = 849.75$, Phi = 0.63, $p < 0.001$)			
White	1199	322	1521
Row percent	79%	21%	100%
Column percent	95%	37%	
Black	62	548	610
Row percent	10%	90%	100%
Column percent	5%	63%	
Total (%)	1290	883	2131
Row percent	59%	41%	100%
Column percent	100%	100%	

Elderly Victim Race Results

Table 9 reveals that elderly victims of different races who were murdered by juvenile White and Black SHOs differed significantly ($\chi^2(1) = 8.43$, $p < 0.01$). White elderly victims were only slightly more likely to be killed by SHOs from their own race (52% vs. 48%). Conversely, all of the Black elderly sexual homicide victims were intra-racially killed ($\chi^2 = 7.2$, df = 1, $p < 0.01$). In terms of the offending pattern against elderly victims, juvenile White SHOs only killed intra-racially. In sharp contrast, 76% of juvenile Black SHOs killed outside their race. There were no cases of White juvenile SHOs having victimized elderly Black females.

When the races of adult sexual murderers are examined, significant differences were found in the killing of elderly victims of different races by White and Black SHOs ($\chi^2(1) = 92.25$, $p < 0.001$). Although most elderly victims murdered by adult SHOs were killed by members of their own race, White victims were less likely to be killed intra-racially (60%) than Black victims (96%). Offender data are again striking. More than half of adult Black SHOs killed inter-racially (56%), when compared to only one of 45 adult White SHOs (2%) who murdered inter-racially.

TABLE 9—*Elderly victim race by juvenile sexual murderer race (*N = *67) and adult sexual murderer race (*N = *394).*

	Offender Race		
Elderly Victim Race	White	Black	Total (%)
Juvenile sexual murderer ($\chi^2(1) = 8.43$, Phi = 0.36, $p < 0.01$)			
White	30	28	58
Row percent	52%	48%	100%
Column percent	100%	76%	
Black*	0	9	9
Row percent	0%	100%	100%
Column percent	0%	24%	
Total (%)	30	37	67
Row percent	45%	55%	100%
Column percent	100%	100%	
Adult sexual murderer ($\chi^2(1) = 92.25$, Phi = 0.48, $p < 0.001$)			
White	180	118	298
Row percent	60%	40%	100%
Column percent	98%	56%	
Black	4	92	96
Row percent	4%	96%	100%
Column percent	2%	44%	
Total (%)	184	210	394
Row percent	47%	53%	100%
Column percent	100%	100%	

*Elderly victims ($\chi^2 = 7.2$, df = 1, $p < 0.01$).

Discussion

This study has described race and age differences in victim–offender relationship for juvenile and adult SHOs using three decades of data from a large national database, the FBI's SHRs. Several limitations of the study will be mentioned here to help frame the ensuing discussion. The limitations of the SHR dataset have been noted earlier, but should be considered in the interpretation of these data. For example, the SHR data form is limited to such basics as offender and victim age, sex, race, victim–offender relationship, type of weapon used, and circumstances (e.g., "bar room brawl," "victim shot by robber"). This level of detail does not allow for a more in-depth investigation into offender factors like the nuances of motivation, presence of paraphilias, degree of psychopathy, and criminal history. Likewise, this dataset provides only limited information for the victims. On the other hand, given that the current study results are derived from national data spanning three decades, we have confidence that the findings are reasonably representative of sexual homicides in the United States for the time span identified. Unfortunately, the number of SHOs of racial backgrounds other than White or Black (specifically American Indian or Alaskan Native, and Asian and Pacific Islanders) accounted for only 1.9% of the total sample. Accordingly, examining crime patterns for offenders of other racial backgrounds in a meaningful way was not possible.

Involvement of White and Black Offenders in Sexual Homicides

According to the U.S. Census Bureau (53,54), Blacks comprised roughly 13% of the U.S. population in 1996 and 2004. Thirty years of arrest data indicated that proportionally speaking, Black offenders in this study of sexual homicide were markedly overrepresented, consistent with Black arrestees being overrepresented in both nonsexual homicides (24) and rapes (33). Although the involvement of Black offenders in sexual homicides was less than their *c.* 50% representation in nonsexual homicides (2), Black offenders' participation in sexual homicide remains a matter of serious concern. Four Black offenders were arrested for every 10 sexual homicides, yet based on demographics one would predict that Black offenders would be involved in about one in eight of these crimes. Similar significant race-based findings were noted when Black juvenile and adult SHOs were studied separately.

In contrast, White offenders were underrepresented based on their proportion of the population. Although Whites comprised 72% of the population in 1996 and 68% in 2004 (53,54), they accounted for only 59% of offenders arrested for sexual homicide. Despite the underrepresentation of Whites proportionately speaking, the modal SHO for this sample was still an adult White offender because there were only two offender groups (Whites, 59%; Blacks, 41%). The modal victim was an adult White female. Caution is advised, however, when reporting these modal categories. Racial disparities are discernible when one notes that in the 3868 sexual homicides identified, although 72% of the victims were White, only 59% of the offenders were White.

Perusal of Tables 6–9 reveals some remarkable race-based disparities in victim–offender relationship. Overall, juvenile and adult White SHOs predominantly killed intra-racially, in the range of 91–100%, across all victim age categories. Of note, there was not one recorded instance of a White juvenile SHO killing a Black elderly victim for the study period, yet 28 crimes involving a Black juvenile SHO killing a White elderly victim were documented.

In contradistinction are the patterns for juvenile and adult Black SHOs; they killed intra-racially or inter-racially in a fashion largely dependent on victim age. The percentages of juvenile Black SHOs who killed intra-racially (range 24–82%) began robustly at 82% for child victims, then progressively dwindled as victim age increased, with 68% killing adolescents intra-racially, 54% killing adults intra-racially, and only 24% killing elderly victims intra-racially. A similar progression, although not quite as marked, of intra-racial killing diminishing with increasing victim age was seen in adult Black SHOs (range 44–82%). They largely killed intra-racially for child victims, also at 82%, then down to 71% for adolescent victims, 63% for adult victims, and only 44% for elderly victims.

In sum, White SHOs were highly likely to kill within their race, "intra-racially" (range 91–100%), whereas Black SHOs killed both intra-racially (range 24–82%) and inter-racially (range 18–76%), with the likelihood of their killing inter-racially increasing as the age of the victim increased. A parallel pattern for elderly victims of SHOs was pointed out by Safarik and colleagues (49): Most of the sexual homicides of elderly victims in their study were White and killed by younger Black offenders.

Reasons for Offender Racial Differences in Sexual Homicide

Why Blacks are overrepresented proportionately speaking in comparison with Whites as SHOs in this study, and why they are more likely to commit inter-racial offenses, are interesting questions that defy easy explanation and raise a number of complex issues for consideration. To mention a few, are Blacks more likely to engage in sexual murder because of some confluence of cultural, socioeconomic, environmental, and geographical influences? Are political factors or racial antagonism somehow at play in a percentage of these crimes?

Do a significantly greater proportion of Black SHOs who set out to commit sexual assault end up killing their victims to diminish the possibility of later being identified to authorities by a surviving victim witness? In a similar vein, might Black SHOs have different arrest histories or more negative experiences with law enforcement than White SHOs that would predispose them to harsher sentences were they arrested for a sexual assault, and thus they would have a greater motivation to kill their victims? Alternatively, could racial stereotypes play into victim response and a greater degree of resistance thereby be mustered by victims of Black SHOs, thereby increasing the likelihood of a fatal outcome given the offender's greater counter-response in turn to achieve and maintain victim control?

Is there a difference in the presence, degree, and expression of psychopathy or sexual sadism in Black versus White SHOs? Might other personality characteristics play a role? What about sexual preferences based on victim race or age? Studies indicate that Black males are more accepting of inter-racial sexual contact than White males (e.g., see [55]). These studies of course refer to consensual, noncriminal interactions, but there is undoubtedly a sexual component at play in nonconsensual sexual crimes as well.

A finding that particularly stood out in these data was that Black juvenile SHOs accounted for all of the Black on Black sexual homicides involving elderly Black victims, and almost half of sexual homicides involving elderly White victims. Are some of the elderly victim crimes accounted for by a malignant gerontophilia, and if so, are there race-based differences in its phenotypic expression? Or, might intended simple burglaries of an aged victim's home at times evolve into a sexual crime based on the opportunity to obtain an easily conquerable sexual victim combined with a polymorphous, psychopathic approach to sexuality?

The statistical rarity of some of the types of sexual homicides might buttress claims for a mental health defense or for mitigation

at sentencing. For example, could the offender have a serious mental disorder that, when combined with alcohol and/or drugs, impaired his ability to distinguish right from wrong or to refrain from sexually acting out in a hypersexualized state?

Addressing such questions as outlined here is far beyond the scope of this study. Clearly, more research is needed to understand the dynamics that propel individuals to commit sexual homicide. In-depth interviews of convicted SHOs are needed to determine whether racial differences in sexual homicides can be explained by sociological or cultural factors, political or criminal justice operational realities, the offenders' life experiences and experiences with criminal justice system, the victims' responses to the offenders during the incident, and/or the diagnosis of mental disorders in the offenders? Many of these explanations could be operationalized and tested. For example, do Black and White SHOs significantly differ from one another in scores on the Psychopathy Check List (PCL-R, 56) overall, by offender age, and/or by victim age type? Do offenders' criminal histories and prior dispositions significantly differ by race? In light of the little research available to date on racial differences among SHOs, directionality of hypotheses seems premature. Null hypotheses, such as the statement that there will be no difference in PCL-R scores between Black and White SHOs, would seem the prudent way to proceed.

Sexual Homicide Victims

Remarkably, Black SHOs (juveniles and adult offenders) were more likely to kill inter-racially against Whites as the age of the victim increased. A large majority of the reported sexual homicide victims in this study were Whites (72%). Although a comparatively small sample, similar findings were found by Myers (4) in a sample of 14 juvenile sexual homicides who primarily killed White victims. This discovery of a predominance of White victims stands in contrast to various studies on nonsexual homicide and sex crime in which it was shown that Black victims were overrepresented (15,16,19,37,38).

Contrary to media depictions and public concern, victims of sexual homicide in this sample were infrequently found to be children (57). In fact, children were the least likely group to be sexually killed (11%), followed by adolescents (12%) and the elderly (13%). Adult victims, accounting for 64% of victims, exceeded all other age categories combined (36%). White SHOs overall were twice as likely to kill children (67% vs. 33%), and both White and Black offenders, whether juveniles or adults, typically killed intra-racially when it came to child victims, with about 95% of Whites and 82% of Blacks doing so.

Implications for Offender Profiling

We believe that the data provided in this article may prove valuable in certain instances by assisting law enforcement with prioritizing their investigation efforts with respect to developing suspects. In general, sexual homicides are intra-racial—most offenders will kill within their own race. For all cases combined in the present study, White offenders killed White victims nearly 80% of the time, and Black offenders killed Black victims in about 90% of cases.

Table 10 portrays what the odds are the killer will be of the same race as the victim based on the victim and offender age. Understandably, investigators in any given case will have varying quality of evidence to suggest the killer's age, ranging from nil to solid. Moreover, many factors will shape the directions of the investigation energy, including crime scene evidence, geographical issues, clues to offender transportation methods. Certainly no formula based on "cold" data such as provided here can speak to an individual crime. However, the ratios provided in Table 10 may prove helpful in certain case scenarios. For example, if a Black adolescent victim is discovered, and eye-witnesses believe the offender of unknown race was an adult, the odds are overwhelmingly suggestive that the offender was Black (49:1).

TABLE 10—*Victim–offender intra-racial/inter-racial odds ratios.**

	Offender Age	
Victim Type	Juvenile	Adult
White child	9:1	12:1
Black child	7:1	8:1
White adolescent	10:1	6:1
Black adolescent	5:1	49:1
White adult	2:1	4:1
Black adult	5:1	9:1
White elderly	1:1	1.5:1
Black elderly	100% of cases were intra-racial	24:1

*Rounded off to nearest whole number except for elderly females.

The intra-racial versus inter-racial odds ratios in Table 10 clearly suggest that as the age of White victims increases, police need to consider the possibility that the SHO may be either White or Black. The likelihood of a juvenile or adult Black offender being involved in the killing of an elderly White female is especially high in relation to other victim age groups. In sharp contrast, if the elderly female is Black, the likelihood that the SHO is White is very remote. The patterns observed here underscore the need for further clinical explanation and law enforcement explanation.

References

1. Wolfgang M, Figlio R, Tracy P, Singer I. The national survey of crime severity. Washington, DC: U.S. Department of Justice, Bureau of Justice Statistics, 1985.
2. Chan HC, Heide KM. Weapons used by juveniles and adult offenders in sexual homicides: an empirical analysis of 29 years of U.S. data. J Investigative Psychol Offender Profiling 2008;5(3):189–208.
3. McNamara JJ, Morton RJ. Frequency of serial sexual homicide victimization in Virginia for a ten-year period. J Forensic Sci 2004;49:1–5.
4. Myers WC. Juvenile sexual homicide. London: Academic Press, 2002.
5. Meloy JR. The nature and dynamics of sexual homicide: an integrative review. Aggress Violent Behav 2000;5:1–32.
6. Chan HC, Heide KM. Sexual homicide: a synthesis of the literature. Trauma Violence Abuse 2009;10(1):31–54.
7. Brownmiller S. Against our will: men, women, and rape. New York, NY: Simon and Schuster, 1975.
8. Burgess AW, Hartman CR, Ressler RK, Douglas JE, McCormack A. Sexual homicide: a motivational model. J Interpers Violence 1986;1: 251–72.
9. Myers WC. Sexual homicide by adolescents. J Am Acad Child Adolesc Psychiatry 1994;33(7):962–9.
10. Revitch E. Sex murder and the potential sex murderer. Dis Nerv Syst 1965;26:640–8.
11. Ressler RK, Burgess AW, Douglas JE. Sexual homicide: patterns and motive. New York, NY: Free Press, 1988.
12. Hill A, Habermann N, Klussman D, Berner W, Briken P. Criminal recidivism in sexual homicide perpetrators. Int J Offender Ther Comp Criminol 2008;52(1):5–20.
13. Heide KM. Weapons used by juveniles and adults to kill parents. Behav Sci Law 1993;11:397–405.
14. Heide KM, Petee TA. Weapons used by juvenile and adult offenders in U.S. parricide cases. J Interpers Violence 2007;22(11):1400–14.
15. Beeghley L. Homicide: a sociological explanation. Lanham, MD: Rowman and Littlefield Publishers, 2003.
16. Hannon L. Race, victim precipitated homicide, and the subculture of violence thesis. Soc Sci J 2004;41:115–21.
17. Maguire K, Pastore A. Sourcebook of criminal justice statistics 2000. Washington, DC: United States Government Printing Office, 2001.

18. Parker KF, Pruitt MV. Poverty, poverty concentration, and homicide. Soc Sci Q 2000;81(2):555–70.
19. U.S. Department of Justice. Sourcebook of criminal justice statistics 1998. Washington, DC: United States Government Printing Office, 1999.
20. U.S. Department of Justice. Sourcebook of criminal justice statistics 1999. Washington, DC: United States Government Printing Office, 2000.
21. U.S. Department of Justice. Crime in the United States 2006, http://www.fbi.gov/ucr/cius2006/offenses/expanded_information/homicide.html (accessed February 24, 2009).
22. U.S. Department of Justice. Crime in the United States 2007, http://www.fbi.gov/ucr/cius2007/offenses/expanded_information/index.html (accessed February 24, 2009).
23. Wilbanks W. Criminal homicide offenders in the U.S.: Black vs. White. In: Hawkins DF, editor. Homicide among Black Americans. Lanham, MD: University Press of America, 1986;43–55.
24. Heide KM, Petee TA. Parricide: an empirical analysis of 24 years of U.S. data. J Interpers Violence 2007;22(11):1382–99.
25. Block R. Homicide in Chicago: a nine-year study (1965–1973). J Crim Law Criminol 1976;66:496–510.
26. Humphrey JA, Palmer S. Race, sex, and criminal homicide offender-victim relationships. In: Hawkins DF, editor. Homicide among Black Americans. Lanham, MD: University Press of America, 1986;57–67.
27. Fox JA, Levin J. Multiple homicide: patterns of serial and mass murder. Crime Justice 1998;23:407–55.
28. Pokorny AD. A comparison of homicide in two cities. J Crim Law Criminol Police Sci 1965;56:479–87.
29. Phillips JA. White, Black, and Latino homicide rates: why the difference? Soc Probl 2002;49(3):349–73.
30. Jacobs D, Wood K. Interracial conflict and interracial homicide: do political and economic rivalries explain white killings of blacks or black killings of whites? Am J Sociol 1999;105(1):157–90.
31. Wadsworth T, Kubrin CE. Structural factors and black interracial homicide: a new examination of the causal process. Criminology 2004;42(3):647–72.
32. Garland T. An overview of sexual assault and sexual assault myths. In: Reddington FP, Kreisel BW, editors. Sexual assault: the victims, the perpetrators, and the criminal justice system. Durham, NC: Carolina Academic Press, 2005;5–27.
33. Belknap J. The invisible woman, 2nd edn. Boulder, CO: Wadsworth, 2001.
34. Palermo GB, Farkas MA. The dilemma of the sexual offender. Springfield, IL: Charles C. Thomas Publisher, 2001.
35. Laufersweiler-Dwyer DL, Dwyer G. Rapists. In: Reddington FB, Kreisel BW, editors. Sexual assault: the victims, the perpetrators, and the criminal justice system. Durham, NC: Carolina Academic Press, 2005;205–31.
36. Bureau of Justice Statistics. Sex offenses and offenders: an analysis of date rape and sexual assault. Washington, DC: U.S. Department of Justice, 1997.
37. Russell DE. Sexual exploitation. Beverly Hills, CA: Sage, 1984.
38. Russell DE. Rape in marriage. Indianapolis, IN: Indiana University Press, 1990.
39. Patton TO, Snyder-Yuly J. Any four black men will do: rape, race, and the ultimate scapegoat. J Black Stud 2007;37(6):859–95.
40. U.S. Department of Justice. Criminal victimization in the United States, 2003 statistical tables, http://www.ojp.usdoj.gov/bjs/pub/pdf/cvus0302.pdf (accessed February 27, 2009).
41. Gacono CB, Meloy JR, Bridges MR. A Rorschach comparison of psychopaths, sexual homicide perpetrators, and nonviolent pedophiles: where angels fear to tread. J Clin Psychol 2000;56:757–77.
42. Grubin D. Sexual murder. Br J Psychiatry 1994;165:624–9.
43. Milsom J, Beech AR, Webster SD. Emotional loneliness in sexual murderers: a qualitative analysis. Sex Abuse 2003;15:285–96.
44. Swigert VL, Farrell RA, Yoels WC. Sexual homicide: social, psychological, and legal aspects. Arch Sex Behav 1976;5(5):391–401.
45. Warren JI, Hazelwood RR, Dietz PE. The sexually sadistic serial killer. J Forensic Sci 1996;41:970–4.
46. Myers WC, Burgess AW, Nelson JA. Criminal and behavioral aspects of juvenile sexual homicide. J Forensic Sci 1998;43(2):340–7.
47. Leach G, Meloy JR. Serial murder of six victims by an African-American male. J Forensic Sci 1999;44(5):1073–8.
48. Safarik ME, Jarvis J, Nussbaum K. Elderly female serial sexual homicide: a limited empirical test of criminal investigative analysis. Homicide Stud 2000;4:294–307.
49. Safarik ME, Jarvis JP, Nussbaum KE. Sexual homicide of elderly females: linking offender characteristics to victim and crime scene attributes. J Interpers Violence 2002;17:500–25.
50. Fox JA, Swatt ML. Uniform crime reports [United States]: supplemental homicide reports with multiple imputation, cumulative files 1976–2005 [Computer file]. Ann Arbor, MI: Inter-university Consortium for Political and Social Research 2008, http://icpsr.umich.edu/ (accessed September 18, 2008).
51. U.S. Census Bureau. Characteristics of children under 18 years by age for the United States, region, states, and Puerto Rico: 2000, http://www.census.gov/population/www/cen2000/briefs/phc-t30/tables/tab01.pdf (accessed March 7, 2009).
52. U.S. Census Bureau. Race and Hispanic or Latino origin by age and sex for the United States: 2000, http://www.census.gov/population/www/cen2000/briefs/phc-t8/tables/phc-t-08.pdf (accessed March 7, 2009).
53. U.S. Census Bureau. Selected social characteristics of the population, by sex, region, and race: March 1996, http://www.census.gov/population/socdemo/race/black/tabs96/tab01-96.txt (accessed February 6, 2009).
54. U.S. Census Bureau. Population by sex and age, for Black alone and White alone, not Hispanic: March 2004, http://www.census.gov/population/socdemo/race/black/ppl-186/tab1.pdf (accessed February 6, 2009).
55. Knox D, Zusman ME, Buffington C, Hemphill G. Interracial dating attitudes among college students. Coll Stud J 2000;34(1):69–71.
56. Hare RD. Psychopathy checklist—revised technical manual, 2nd edn. Toronto: Multihealth Systems, Inc., 2003.
57. Heide KM, Beauregard E, Myers W. Sexually motivated child abduction murders: synthesis of the literature. Vict Offenders 2009;4(2):58–75.

Additional information and reprint requests:
Kathleen M. Heide, Ph.D.
University of South Florida
4202 East Fowler Avenue
Social Sciences Room 311
Tampa, FL 33620-8100
E-mail: kheide@cas.usf.edu

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01456.x
Available online at: interscience.wiley.com

PAPER

QUESTIONED DOCUMENTS

Carolyne Bird,[1,2] *B.Sc. (Hons); Bryan Found,*[1,3] *Ph.D.; and Doug Rogers,*[1] *Ph.D.*

Forensic Document Examiners' Skill in Distinguishing Between Natural and Disguised Handwriting Behaviors

ABSTRACT: Disguised handwriting is problematic for forensic document examiners (FDEs) and attracts higher misleading and inconclusive rates on authorship opinions than does genuine writing (Found B, Rogers D, International Graphonomics Society, 2005). There are currently no published empirical data on FDEs' expertise in distinguishing between natural and disguised writing behavior. This paper reports on the skill of FDEs for determining the writing process of 140 pairs of natural and disguised writings and compares their results with those of a control group of laypeople. A significant difference was found between the examiner and lay group. FDEs' expertise is characterized by their conservatism, where FDEs express a higher proportion of inconclusive opinions (23.1% for FDEs compared to 8.4% for the control group). This leads to the FDEs expressing a smaller percentage of misleading responses when calling writings as either naturally written or disguised (4.3% for FDEs compared with 12.2% for the control group).

KEYWORDS: forensic science, questioned document examination, forensic handwriting examination, handwriting, validation, disguise

Handwriting identification evidence has been accepted in courts for over 100 years (1). However, unlike other forensic science disciplines such as DNA evidence, handwriting identification evidence did not grow out of an academic field of research, but from a need of the court. For many years, there was no perceived need for validation of the skill of forensic document examiners (FDEs), and little was carried out in this area until the publication of a series of articles (2,3) questioning FDEs' claims of scientific expertise propelled research into validating their purported skill. In the past 15 years, a number of studies (4–9) have compared examiners' opinions with those of laypeople and found that FDEs do have expertise in relation to opinions on questioned handwriting and signatures. This expertise is characterized by the conservatism of FDEs, which results in significantly lower misleading (erroneous) opinions when compared to laypeople. However, in none of the reported expertise studies were participants required to provide authorship opinions on unnatural questioned writings that were a mixture of disguised handwritten text by the genuine writer and simulated writings by someone other than the genuine writer. Recent research involving only FDEs as participants has identified disguised and simulated handwriting types as problem areas for FDEs when the questioned writings have included a mixture of disguised and simulated writings (10,11). The studies have shown that these writing types attract higher incorrect and inconclusive rates on authorship opinions than genuine writing does.

Given that FDEs have been shown to have a higher error rate for unnatural handwriting than normal writing, if FDEs can identify a piece of questioned writing as the product of an unnatural (disguised or simulated) writing process, they can express this "first-stage" opinion without necessarily proceeding on to express an opinion on authorship (a "second-stage" opinion). Conservatism with regard to second-stage opinions should help to reduce any potential errors in casework. However, there are currently no published empirical data on FDEs' capacity to distinguish between natural and disguised writing behavior in a blinded fashion. In this paper, we attempt to examine whether FDEs have a skill in determining which of a pair of handwriting samples was written using a disguise process. To do this, their opinions on the process of production of 140 pairs of natural and disguised handwriting samples were examined and compared to the opinions of a control group of laypeople.

Methods

Participants

Results reported here are from two groups of participants (forensic handwriting experts and laypeople) who provided independent opinions on the process of production of the supplied handwriting samples.

Eleven FDEs from two countries took part in the trial, six from the United States and five from Australia. All participants have had a minimum of two years full-time government laboratory training

[1]Handwriting Analysis and Research Laboratory, School of Human Biosciences, La Trobe University, Bundoora, Victoria 3086, Australia.
[2]Document Examination Section, Chemistry and Materials Group, Forensic Science SA, 21 Divett Place, Adelaide, South Australia 5000, Australia.
[3]Document Examination Unit, Digital and Document Evidence Branch, Victoria Police Forensic Services Centre, Macleod, Victoria 3085, Australia.

Received 11 May 2009; and in revised form 16 July 2009; accepted 15 Aug. 2009.

under a qualified examiner and have worked in government laboratories. Their training meets the standard set out in the ASTM E2388-05 Standard Guide for Minimum Training Requirements for Forensic Document Examiners. All of the participants have been authorized by their employers to release opinions regarding the authorship of questioned handwriting and signatures. At the time of data collection, 10 FDEs were working in government laboratories and one was working privately

None of the 10 laypeople had training in forensic handwriting and signature examinations; however, two reported having some knowledge of the discipline. The self-reported knowledge of forensic handwriting examination came from a participant who works within a forensic science laboratory in an unrelated discipline and a student who had undertaken a short work experience placement within a forensic handwriting examination laboratory.

The authors are aware that the number of participating FDEs is small. However, with the relatively small number of FDEs worldwide available (and willing) to be tested and this trial comprising one-third of a larger project requiring the involvement of FDEs, a significantly larger sample size could not practically be obtained.

Materials

The trial consisted of 140 pairs of handwritten text, each pair written by one writer, with 70 writers in total. One of the samples in the handwriting pairs was the writer's normal handwriting, in either upper or lowercase print, and the other was a disguised handwriting sample in the same case. No directions were given to the handwriting providers as to how to disguise their handwriting. All writings were made using the same make of ballpoint pen and the same make of white paper. The samples were scanned at 600 dpi, and inkjet printed into a booklet and converted into high-resolution PDF files.

Procedure

The FDE group was provided with a sample booklet of inkjet-printed images of the handwriting samples, while the laypeople were provided with a CD containing PDF files of the samples. The different mode of application for the trials is a result of the large number of samples and time constraints involved in distributing a single sample folder to all participants. The authors note that the different form of material examined by laypeople and FDEs is a limitation of this study. However, anecdotal evidence indicates that FDEs are equally effective examining digital representations and hardcopies, and it could be argued that digital files allow easier magnification of the fine features compared with a printed copy of the original.

Every participant was given an answer-recording booklet and informed that each sample pair was written by one writer and consisted of one naturally written and one disguised handwriting sample. They were asked to make an assessment as to which of the sample pair (A or B) was disguised by entering the appropriate code (A, B, or I for an inconclusive opinion) in the answer booklet comprising boxes corresponding to each one of the questioned sample pairs. If an inconclusive opinion was given, the participant was required to record a forced opinion (A or B) for the handwriting pair. Only the unforced opinions of the participants are included in this paper. In addition, the disguise strategy or strategies the participant believed the writer used was recorded in the answer booklet by marking the appropriate box(es) corresponding to writer-reported disguise strategies and those commonly mentioned in the literature (12–17). An analysis of this aspect of the trial will be reported separately.

Ethics Approval

Approval for this study was obtained from the La Trobe University Human Ethics Committee on the basis that the handwriting providers and test participants gave full consent for samples of their handwriting, or research data provided by them respectively, to be included in published material, on the condition that neither their name nor any other identifying information be used.

Analysis

Participants' responses were marked as correct, misleading, or inconclusive and analyzed to produce scores for the groups of FDEs and laypeople. To determine whether there were differences in scores between the examiner group and the lay group, a series of planned comparisons were undertaken using independent sample *t*-tests (two-tail).

Results

When the overall results for the FDEs and laypeople are compared, it is clear that there is a difference in the response profiles between the two groups. Table 1 provides a summary of the correct, misleading, and inconclusive responses for the group of FDEs and laypeople tested.

Previous studies (4–9) have found that when compared to laypeople, FDEs do have a greater skill in relation to expressing authorship opinions on questioned handwriting and signature samples. The results reported in this paper are aligned with the trend evident in these previous findings where FDEs exhibit a greater skill in determining which of a pair of handwriting samples was disguised. Furthermore, as in the research reported by Sita, et al. (7), FDEs were found to be more conservative than their untrained counterparts, with an inconclusive rate more than twice that of laypeople (23.05% versus 8.43%, $t = 3.39$, df = 19, $p = 0.003$). Although the raw scores indicate that laypeople have a higher correct rate than FDEs, this difference is not significant ($t = 1.81$, df = 19, $p = 0.086$); however, the error rate of FDEs is less than a third of the error rate of laypeople (3.38% compared with 11.43%, $t = 4.08$, df = 19, $p < 0.001$). This means that when participants were willing to give an opinion on which of a pair of handwriting samples was disguised, the FDEs have a significantly higher correct rate than laypeople (95.66% correct called compared to 87.84%, $t = 3.86$, df = 19, $p = 0.001$). This is where the skill of FDEs rests: when an opinion is given, it is more likely to be correct than when a layperson gives an opinion. Table 2 gives the correct and misleading called rates for the two groups.

The high inconclusive rate for FDEs may be the result of unexpected similarities in the handwriting sample pairs, where either the writer is not very adept at introducing or maintaining a disguise, so the two samples appear very similar; or where the natural writing of the writer is messy, internally inconsistent, and/or lacks fluency,

TABLE 1—*Scores of grouped responses for FDEs and laypeople.*

Score	FDEs	Laypeople
% Correct	73.38	80.07
% Misleading	3.38	11.43
% Inconclusive	23.05	8.43

TABLE 2—*Scores of grouped called responses (excluding inconclusive responses) for FDEs and laypeople.*

Score	FDEs	Laypeople
% Correct Called *	95.66	87.84
% Misleading Called	4.34	12.16

*#Correct*100 / (#Correct + #Misleading).

which are some of the traditional hallmarks of unnatural writing (12–18). In these cases where the disguised sample is not clear, the FDE, having been trained to be conservative, will decline to opine as to which sample is disguised, while a layperson is more likely to hazard a guess. An example of a sample pair with an inadequate disguise strategy is given in Fig. 1. In this case, 6 of the 11 FDEs gave an inconclusive opinion, while the remaining examiners correctly identified the disguised sample. Conversely, only 2 of the 10 laypeople declined to offer an opinion, with 6 correctly identifying the disguised sample and the remaining 2 providing an erroneous opinion.

The second scenario is illustrated in Fig. 2, where both the writer's normal and disguised handwriting samples appear internally inconsistent, with some tremor in the line trace. Here, 8 of the 11 FDEs preferred not to opine as to which was the disguised sample and gave an inconclusive opinion. The three FDEs who did offer an opinion were correct. The laypeople had a 70% correct rate for this sample pair; however, only one layperson offered an inconclusive opinion, and two were incorrect.

If we examine the individual responses of participants in the two groups (Fig. 3), the trend touched on above becomes apparent, and distinct examiner and laypeople profiles emerge. The typical examiner profile shows an inconclusive rate of between 14.3–37.9%, with a misleading rate of 1.4–4.3%, while the typical layperson profile exhibits an inconclusive rate between 0 and 14.9% and misleading rate of 6.4–20.7%. Only two of the laypeople (L04 and L15) exhibit the "examiner profile," and one examiner (E08) displays a "layperson profile." The laypeople with examiner-like profiles were the only two laypeople who self-reported knowledge of forensic handwriting examination at some level.

In line with previous research (7,19), there is no correlation between FDEs' years of experience and their skill for determining which of a pair of handwriting samples was written using a disguise process. Regression statistics on the % correct for examiners were $R = 0.32$, $p = 0.340$ and for % correct called, $R = 0.07$, $p = 0.836$.

As intuitively expected, many disguised handwriting samples are easily identified by their stilted, unnatural appearance (Fig. 4). These types of sample pairs did not pose a problem for either group.

However, other writers were able to employ a reasonably consistent, fluently executed disguise, which both FDEs and laypeople had trouble identifying. The sample pair given in Fig. 5 attracted high erroneous opinions across both groups, with all of the examiners who offered an opinion (72.7% of examiners) identifying the naturally written sample as disguised. The remaining three FDEs gave an inconclusive opinion. The laypeople faired little better, with a 90% error rate and one person correctly identifying the disguised sample.

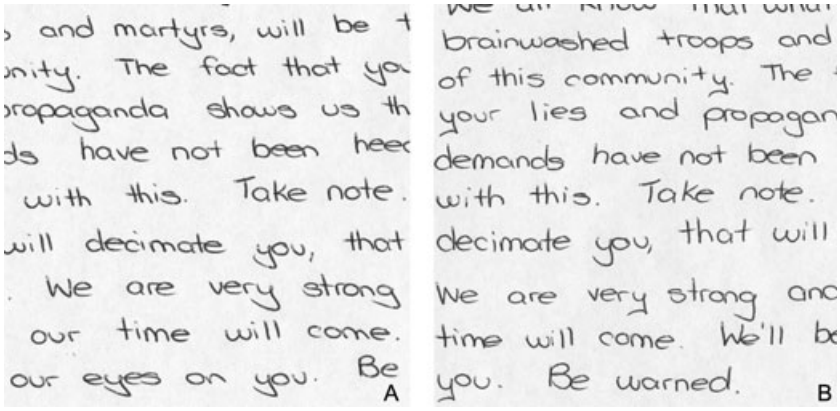


FIG. 1—*Portion of a handwriting sample pair written by one person, where the disguised sample was not clearly identified by some FDEs due to a subtle disguise strategy. (A) Normal handwriting sample. (B) Disguised handwriting sample.*

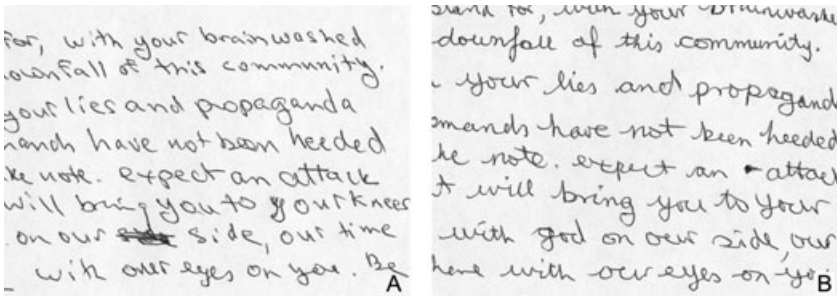


FIG. 2—*Portion of a handwriting sample pair written by one person, where the disguised sample was not clearly identified by some FDEs due to a lack of fluency in the normal writing. (A) Normal handwriting sample. (B) Disguised handwriting sample.*

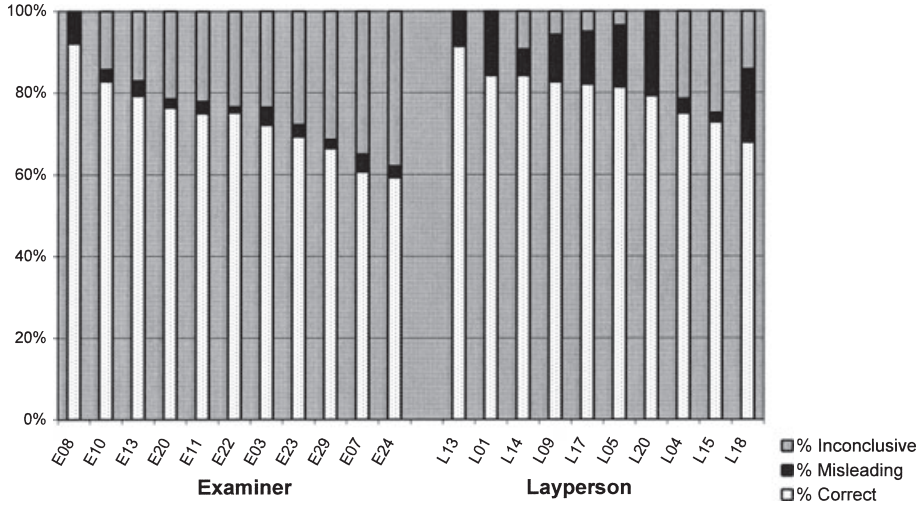


FIG. 3—*Individual participants' correct, misleading, and inconclusive responses for handwriting sample pairs. Note: E denotes a forensic handwriting examiner, L denotes a layperson.*

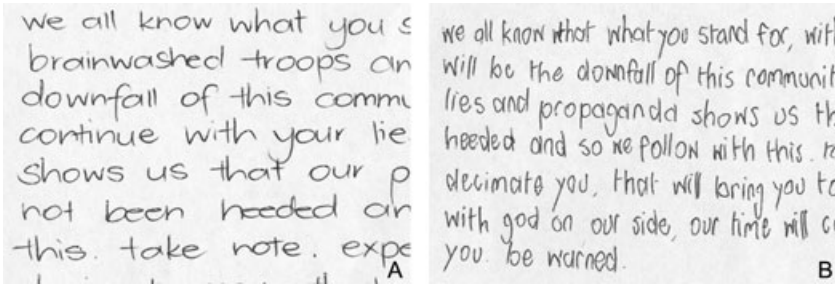


FIG. 4—*Portion of a handwriting sample pair written by one person, where the disguised sample was clearly identified by participants. (A) Normal handwriting sample. (B) Disguised handwriting sample.*

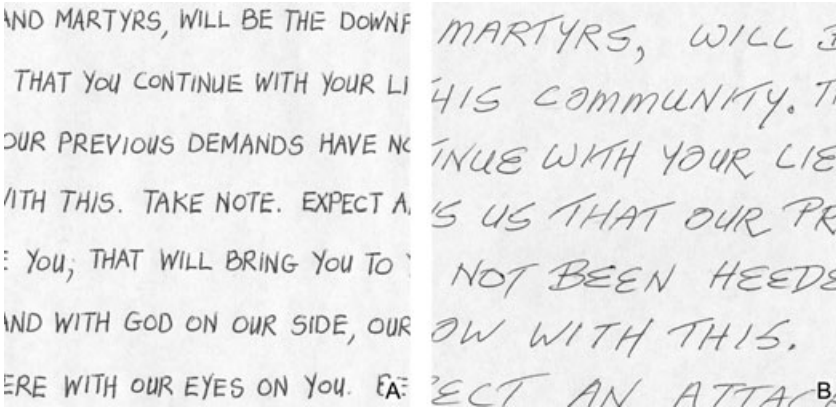


FIG. 5—*Portion of a handwriting sample pair written by one person, where the disguised sample was thought by some FDEs to be naturally written. (A) Normal handwriting sample. (B) Disguised handwriting sample.*

Interestingly, there were a few sample pairs where FDEs had no problem or uncertainty in identifying the disguised sample, whereas laypeople attracted an error. Figure 6 is one such example where FDEs had a 100% correct rate, while laypeople had a 30% error rate. Further investigation of the disguise strategies participants believed were employed in creating these samples may help to identify predictors of disguised handwriting, and further characterize the skill of FDEs undertaking this task.

Discussion

As a group, forensic document examiners do possess a skill in determining which of a pair of handwriting samples, written by one writer, is disguised. The reported results show that FDEs have what would be considered acceptably low misleading responses (≤4.3%). On a practical level, this skill is used by FDEs routinely in casework examination when assessing the likelihood of a sample

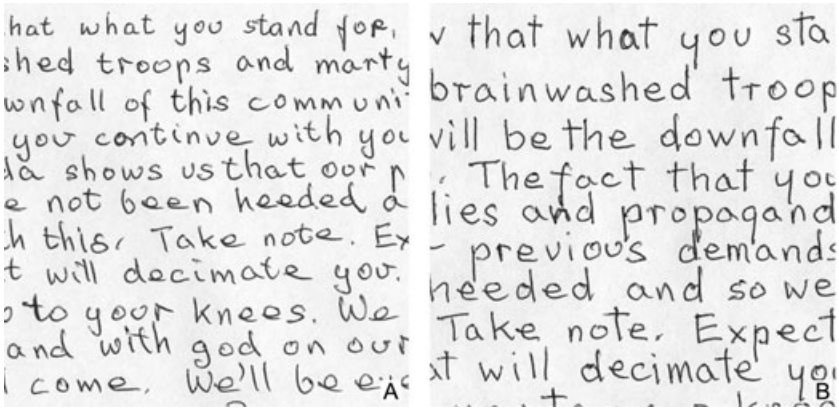


FIG. 6—*Portion of a handwriting sample pair written by one person, where the disguised sample posed no problem to FDEs but attracted an error from laypeople. (A) Normal handwriting sample. (B) Disguised handwriting sample.*

of questioned handwriting, or in some instances the comparison handwriting, being the product of disguise behavior. If a comparison sample is considered to be the result of an unnatural writing process, further comparison material may be required. Alternately, if a questioned handwriting sample is opined to be unnaturally written, this should lead the FDE to carefully consider whether any authorship opinion should be formed given the reported high misleading scores associated with this category of questioned writing (10). Even without an opinion regarding authorship, the first-stage process opinion that a sample of handwriting is something other than genuine may be of strong evidential value.

The high inconclusive rate of FDEs may be due in part to the presence of minimal observed differences between natural and disguised samples. For example, when a writer is poor at introducing or maintaining a disguise, there will appear to be many similarities between the two samples, generally causing FDEs to decline to offer an opinion on which of the pair is disguised. If the writer's natural handwriting is not highly skilled, it may exhibit the hallmarks of unnatural writing, which we would also expect to see in the disguised handwriting sample. The lack of a clear difference in fluency causes the FDE to be conservative. In other cases, the handwriting providers were able to successfully create a disguised sample that appeared fluently written and different from their natural writing. The combination of comparable fluency but different features and forms also brings about conservatism in FDEs.

Acknowledgments

We are grateful to the handwriting providers, forensic document examiners, and laypeople who generously gave their time to participate in the study.

References

1. Huber RA, Headrick AM. Handwriting identification: facts and fundamentals. Boca Raton, FL: CRC Press, 1999.
2. Risinger DM, Denbeaux MP, Saks MJ. Exorcism of ignorance as a proxy for rational knowledge: The lessons of handwriting identification "expertise". U Pa L Rev 1989;137:731–92.
3. Risinger DM, Saks MJ. Science and nonscience in the courts: Daubert meets handwriting identification expertise. Iowa Law Rev 1996;82:21–74.
4. Found B, Sita J, Rogers D. The development of a program for characterising forensic handwriting examiners' expertise: signature examination pilot study. J Forensic Doc Exam 1999;12:69–80.
5. Kam M, Wetstein J, Conn R. Proficiency of professional document examiners in writer identification. J Forensic Sci 1994;39:5–14.
6. Kam M, Fielding G, Conn R. Writer identification by professional document examiners. J Forensic Sci 1997;42:778–86.
7. Sita J, Found B, Rogers D. Forensic handwriting examiners' expertise for signature comparison. J Forensic Sci 2002;47:1117–24.
8. Kam M, Gummadidala K, Fielding G, Conn R. Signature authentification by forensic document examiners. J Forensic Sci 2001;46:884–8.
9. Dyer AG, Found B, Rogers D. Visual attention and expertise for forensic signature analysis. J Forensic Sci 2006;51:1397–404.
10. Found B, Rogers D. Problem types of questioned handwritten text for forensic document examiners. Proceedings of the 12th Conference of the International Graphonomics Society; 2005 June 26 – 29; Salerno, Italy. International Graphonomics Society, 2005;8–12, http://www.graphonomics.org/publications.php
11. Found B, Rogers D. The probative character of forensic handwriting examiners' identification and elimination opinions on questioned signatures. Forensic Sci Int 2008;178:54–60.
12. Alford E. Disguised handwriting. J Forensic Sci 1970;15:476–88.
13. Keckler JA. Felonious disguise. Int J Forensic Doc Exam 1997;3:154–8.
14. Konstantinidis S. Disguised handwriting. J Forensic Sci Soc 1987;27:383–92.
15. Harrison WR. Disguise. In: Suspect documents: their scientific examination. London: Sweet & Maxwell, 1958;349–72.
16. Regent J. Changing slant: is it the only change? J Forensic Sci 1977;22:216–21.
17. Totty RN. Recent developments in handwriting examination. In: Maehly A, Williams RL, editors. Forensic science progress, Vol. Berlin: Springer, 1991;5.
18. Muehlberger RJ. Identifying simulations: practical considerations. J Forensic Sci 1990;35:368–74.
19. Found B, Rogers D, Herkt A. The skill of a group of forensic document examiners in expressing handwriting and signature authorship and production process opinions. J Forensic Doc Exam 2001;14:15–30.

Additional information and reprint requests:
Carolyne Bird, B.Sc. (Hons)
Forensic Science SA
21 Divett Place
Adelaide, South Australia 5000
Australia
E-mail: carolyne.bird@sa.gov.au

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01449.x
Available online at: interscience.wiley.com

PAPER

QUESTIONED DOCUMENTS

Susan J. Turnbull,[1,2] *M.A., D.Clin.Psy., M.Sc.; Allison E. Jones,*[1] *B.Sc., M.Sc., Ph.D.; and Mike Allen,*[1] *B.A., M.Sc.*

Identification of the Class Characteristics in the Handwriting of Polish People Writing in English

ABSTRACT: An investigation was carried out to identify the class characteristics of Polish people writing in English and to specifically identify those characteristics that separate Polish handwriting from English handwriting. In the first stage, 40 Polish and 40 English handwriting samples were collected and systematically examined. In total, 31 features were identified that occurred in ≥25% of the Polish handwriting samples and therefore considered class characteristics. Of these, chi-square analyses identified 21 class characteristics that occurred significantly more in Polish compared to English handwriting. Twenty-one of the class characteristics in the Polish handwriting had similar constructions to the copybook pattern thus supporting the theory that class characteristics frequently stem from the taught writing system. In the second stage, an algorithm was developed using seventeen of the class characteristics that successfully discriminated between a further 13 Polish and 12 English handwriting samples.

KEYWORDS: forensic science, questioned documents, class characteristics, Polish handwriting, English handwriting, handwriting identification

This study was designed to identify the class characteristics in Polish handwriting. In the first stage, it identified letter shapes that are common and occur frequently in the English handwriting of Polish people compared to that of English people. The number of class characteristics in Polish handwriting that derive from the taught copybook style was also quantified. In the second stage, an algorithm was developed from the identified class characteristics and tested in its ability to classify whether unknown pieces of writing were written by Polish or English individuals.

The ability to distinguish between class characteristics and individual characteristics is widely believed to be one of the fundamental bases by which a document examiner can make judgments and form opinions to distinguish one person's writing from another (1,2). This knowledge can be built up through the experience of studying the handwritings of a population but it can also be aided by consulting reference libraries and literature that report on particular characteristics found in handwriting by people taught in certain styles and geographical areas (3–6). It has been argued that a document examiner's unfamiliarity with the class features of a given population can lead to the misidentification of a feature as unusual when it is in fact a common feature. This can lead to an error of judgment that compromises the handwriting examination (7). It is important, therefore, to build up a body of literature to which an examiner can refer to when presented with a rarely encountered script.

[1]School of Forensic and Investigative Sciences, University of Central Lancashire, Preston PR1 2HE, U.K.
[2]Forensic Handwriting and Document Investigation, 25 Cameron Street, Stonehaven AB39 2BL, U.K.

Received 20 Mar. 2009; and in revised form 12 July 2009; accepted 19 July 2009.

Several studies have addressed the influence of the taught writing style in the class characteristics of adult handwriting. When this has been empirically tested, contradictory findings have emerged. This is perhaps because of the differing methodologies employed in studies, for example the parameters set as to what is an adherence and what is a departure from a copybook style; or the method of establishing exposure to a particular copybook in childhood. For example, Muelberger et al. (8) looked at the formation of "th" in the writings of 200 individuals who were taught to write in schools using the Palmer copybook. It was shown that although only 1% of people continued to adhere strictly to the Palmer standard, in that every feature of the letter combination conformed to the copybook construction, 78% of people retained at least one feature of the copybook standard.

More recently, Simner and Smits-Englesman (9) found that students who had been taught to write in Canada and the Netherlands were more likely to have features in their writing that reflect letter shapes found in copybook patterns from their respective countries. This lends some support to the importance of having knowledge of the national copybook systems in that they demonstrated that the influence does persist in adult writings, and that reference to copybook catalogues may be useful in identifying class characteristics when an examiner is confronted with a foreign script that he is unfamiliar with. However, the authors also found that within the letter patterns used, both the Dutch and Canadian students frequently used formations that differed from the copybook. They estimated that 70–80% of the Dutch and Canadian participants, respectively, had been exposed to the copybook patterns that they used for comparison during their schooling. This implies that a sizeable minority of their sample had been exposed to alternative or no copybooks during their schooling, which may explain the frequent use of different letter patterns. For example, in the

Netherlands, it has been estimated that at least 10 new copybooks have been created since the 1960s (Fagel, Netherlands Forensic Institute, 2009, personal communication).

Attempts at developing methods of identifying the nationality of the writers of pieces of writing of unknown provenance using copybooks have, however, been limited in their success. Cha et al. (10) reported on their development of a copybook database that used a computerized pattern-matching algorithm as a means of identifying the country of origin of a writer. Their system was only able to correctly match four of 15 handwritten documents to their writers' countries of origin. The difficulty in identifying writing in this way may be because of the wide variability in the handwriting characteristics of natives of some countries. In particular, they found that handwriting from England was most likely to be confused with styles from other countries. This may be because of the lack of emphasis in England on any particular copybook system in the teaching of handwriting (11).

The style of writing taught in Poland appears to have remained consistent over the years. Copybooks obtained from 1974 and 1975 from a database held by the European Network of Forensic Science Institutes (12) are identical to the ones circulated in Poland in 2008 (13,14). This apparent homogeneity in the taught writing style suggests that it is possible that the class characteristics in Polish writing may be strongly influenced by the copybook.

Class characteristics have also been identified in groups of writers without reference to the taught copybook styles. The validity of this approach is supported by the observation that similar frequencies for certain letters in the handwriting of population groups have been recorded by separate studies. For example, Livingston (15), Horton (16), and Zimmerman (17) found comparable frequencies of the Greek "ε" (3%, 1.2%, and 2.7%); and Horton (16) and Zimmerman (17) reported similar frequencies of the reverse bowl or printed "b" (16.0% and 18.5%) in U.S. populations. Tull (18) systematically identified the frequencies of different number forms in a U.S. population using a predetermined classification system and also presented previously unpublished data from two studies completed by others in the United States (U.S.) and in England. The number forms that were most frequently observed were similar in all three studies, highlighting that there can be considerable overlap in common characteristics between different groups of writers.

It can be argued that for a document examiner it is most pertinent to be aware of class characteristics in population subgroups that occur commonly in subgroups and do not occur frequently in the larger group. These features can then be used as a means of distinguishing that group as well as highlighting features that an examiner, otherwise unfamiliar with the subgroup, may mistakenly think of as individual rather than as class characteristics (7). The majority of these studies have been conducted in immigrant populations to the U.S. (6,19–21). One further recent study has addressed the characteristics in different ethnic groups in Singapore (5). This study draws on the research methods used in three of these studies in order to develop an objective, statistically valid means of identifying class characteristics in an immigrant population to the United Kingdom (U.K.).

Berthold and Wooton (19) initially selected the features to look for in their samples of writing by noting 27 letter features that examiners perceived to be unusual in immigration forms compared to the writing of U.S.-educated citizens. They then examined 50 forms from immigrants from each of nine different countries in order to assess the frequency of occurrence of these hypothesized features. They used cut-off points to identify features that they considered as uncommon (0–10%), fairly common (11–24%), common (25–39%), and very common (40+%) in samples from each country. In this way, it was possible to compare the prevalence of a feature in a population between the different country groups.

King (21) also used an objective approach in her examination of the writings of Mexican immigrants to the U.S. Fifty hand-printed exemplars from financial crime cases were examined and each different construction of a letter appearing in an individual's writing was cut and pasted into letter groups. The entire dataset was then examined, letter by letter, and the number of people that used each different identified letter construction once or more in their writing was identified. She identified a letter construction as a class characteristic if it was used by 10% or more of the writers. This approach removes any preconceived opinions regarding which common letter constructions are to be counted, and hence allows a more objective means of measuring class characteristics.

Cheng et al. (5) examined the handwritings of the three main ethnic groups in Singapore: Chinese, Malay, and Indian. Each group, as well as writing in English at school, learn their own mother language and it was hypothesized that the influence of the other language, and its associated writing system, would be observed in the class characteristics present in their English writings. They gathered samples from around 50 individuals from each population subgroup, observed the features, and selected 10 features that occurred with the highest frequency. They then compared the frequencies that each feature occurred between the groups using a two-by-three chi-square analysis to identify whether these features occurred significantly more in some groups than others. This was followed by a two-by-two chi-square analysis to determine in which particular group the feature was significantly more common. In this way, they identified six features that occurred commonly and could be ascribed uniquely to either the Chinese, Malay, or Indian group. This study was commendable in that it managed to statistically discriminate between features that were more common in one group than another.

The authors are not aware of any published study specifically designed to identify the class characteristics in the handwriting of adults native to Poland or any studies that have addressed the links between the Polish copybook standard and the handwriting of Polish adults.

Methods

The two stages of the study required similar information to be collected for analysis. Therefore, the participants and procedures for collecting handwriting samples are first described before the methods of analysis for the two separate stages are detailed.

Participants

Samples of handwriting were obtained from 53 Polish nationals, taught to write in Poland, and 52 control samples from English nationals, taught to write in English in England. One sample was also obtained from a Slovakian national taught to write in Slovakia. All participants were residents of the U.K. Ethical approval for the study was given by the University of Central Lancashire Ethics Committee.

Procedure

Participants completed a demographic form detailing their age, gender, handedness, level of education, and how long they have lived in the U.K. Participants were required to copy a typed template onto lined paper in their normal handwriting using a ballpoint pen. The template included the "London Letter" (5) in normal

lower case script, "THE QUICK BROWN FOX JUMPS OVER THE LAZY DOG" in block capitals and the numbers 0–9 (twice). This has been recommended as a method for collecting a comprehensive handwriting sample (22).

Identification of Class Characteristics

Forty handwriting samples each from the Polish and English participants were used to identify the class characteristics. The handwriting samples were examined blind, concealing the country of origin of each participant from the investigator. All handwriting samples were scanned at 600 dpi using an Epson® Stylus DX8400 scanner (Seiko Epson Corporation, Suwa, Nagano, Japan) and viewed at magnifications of up to 14:1 using imaging software (Jasc® Paint Shop Pro 7, Corel Corporation, Ottowa, Ontario, Canada) to allow detailed examination of the writings. Characteristics were identified by two examinations that are described in further detail later. If a characteristic appeared in ≥25% of the writings of the Polish participants, it was accepted as a class characteristic. This is the threshold defined by Berthold and Wooton (19) above which they considered characteristics to be "common" or "very common" and is higher than that used by King (21) who used a 10% cut-off point.

Examination One: Suggested Characteristics

The handwriting samples were examined for the frequency of features previously suggested as possible class characteristics by an experienced handwriting examiner (Hughes, Document Evidence Ltd, 2008, personal communication). Each of the 80 samples was examined blind to nationality to see whether the following features suggested by an experienced examiner were present or not: the Z and J with loops at the corners; the 1 with the introductory stroke; the 9 formation; the crossed 7; the M, N, and A starting with an upstroke; the 2 stroke P, R, and B; and the shape of the M (Hughes, Document Evidence Ltd, 2008, personal communication). These features are illustrated in Fig. 1 in the results. The number of Polish and English individuals with the hypothesized construction occurring once or more in their handwriting sample was counted and the frequency that the feature occurred calculated.

Examination Two: Exploration for Characteristics

A methodical exploration of all the letter constructions in a subset of 15 of the Polish samples was completed to identify further features that may be class characteristics before examining the entire 80 samples for the frequency of their presence or absence. After randomly selecting a subset of 15 Polish writing samples, the construction of each upper and lower case character was studied. The handwriting samples were analyzed using a methodology similar to that employed by King (21) in her study of Mexican handwriting: different characteristics in each writing sample were digitally cut and pasted (using Jasc® Paint Shop Pro 7 and Microsoft® Word) into separate character groups for further study. It was then possible to identify features that occurred in three or more of the subset of 15 samples and considered by the investigator to be distinctive. Following this, all 80 of the Polish and English samples were then examined blind to nationality for the presence or absence of each of these features. It is acknowledged, however, that the blinding may have been compromised as the 15 samples that had been examined in detail earlier may have been recognized by the investigator. The total number of individuals with each feature occurring once or more in their handwriting sample was recorded and the frequency of the characteristic occurring calculated separately for the Polish and English samples.

Statistical Analysis

SPSS version 16 (SPSS Inc., IBM, Chicago, IL) was used for the data analysis.

Comparison with Copybook

Polish copybooks were obtained from the database of copybooks held by the European Network of Forensic Science Institutes (12), and contemporary copybooks were purchased in Poland (13,14). All the features found to occur frequently in the Polish sample were compared with the Polish copybook patterns to identify which class characteristics were consistent with the copybook and which ones differed from the copybook pattern. The number of identified class characteristics derived from the copybook and present in each sample of writing was calculated.

Development of Algorithms

The features that were identified as significantly distinguishing between Polish and English handwriting were ranked according to the chi-square values. Features that were found to occur in ≥11% of the English samples were removed from the algorithms as this is the cut-off defined as fairly common by Berthold and Wooton (19). Three algorithms were developed to be tested using the results of the identification of the class characteristics in the first 80 handwriting samples:

- Algorithm 1: One point for each feature present in a sample.
- Algorithm 2: Two points for features that occurred in ≥75% of the Polish writings and one point for features occurring in 25–74% of the samples.
- Algorithm 3: Three points for features occurring in ≥75% of the Polish writings, two points for features occurring in 50–74% of the samples, and one point for features occurring in 25–49% of the samples.

Cut-off points were developed for the algorithms as follows. The maximum English score was taken as the cut-off below or equal which to indicate a non-Polish sample; the 10th percentile score from the Polish samples was taken to be the cut-off above or equal which to indicate a Polish sample; and the scores in between these two points were taken as being in the area of uncertainty.

Testing the Algorithms

The remaining samples from 13 Polish, 12 English, and one Slovakian writer were used to test the algorithm. The results from Stage One were used to identify the features that satisfied the criteria for use in the algorithm. The samples were then examined for these features with the investigator blind to the nationality. Scores were assigned for each algorithm to see whether the algorithms could successfully distinguish between the samples and which one, if any, was best at identifying the Polish participants.

Results

A total of 106 samples were collected for the study. It was intended that all samples should be in ballpoint pen. The majority of the handwriting samples were completed in ballpoint pen. Three Polish samples and three English samples were written in roller-

TABLE 1—*Participant demographics.*

		Polish (n = 40)	English (n = 40)	Statistics
Age	Median (25–75‰)	26 years (24.25–30.75)	28 years (23.00–31.75)	Mann–Whitney U = 744.5, p = 0.592
	Range	19–51	17–42	
Gender	Male	13	15	χ^2 = 0.220, df = 1, p = 0.639
	Female	27	25	
Handedness	Right	38	37	Fisher's Exact Test p = 1
	Left	2	3	
Education	No exams	2	0	χ^2 likelihood value = 6.6, df = 4, p = 0.158
	School exams	11	8	
	Vocational	4	1	
	Undergraduate	17	21	
	Postgraduate	6	10	
Time in U.K.	Median (25–75‰)	2 years (1.5–3)	n/a	
	Range	0.08–10		
Age moved to U.K.	Median (25–75‰)	24 years (22–27)	n/a	
	Range	17–49		

ballpen and one Polish sample was written in pencil. The type of writing instrument did not, however, impair the identification of the class characteristics.

Identification of Class Characteristics

Eighty samples (40 Polish and 40 English) were examined in this stage. Table 1 provides the demographic details of the participants. There was no significant difference between the Polish and English samples in terms of age (Mann–Whitney U = 744.5, p = 0.592), gender (χ^2 = 0.220, df = 1, p = 0.639), handedness (Fisher's exact test, p = 1), or level of education (χ^2 likelihood value = 6.6, df = 4, p = 0.158). The Polish participants had resided in the U.K. for an average of 2 years with a range from 1 month to 10 years. All Polish participants were taught to write in Poland and were at least 17 years old on moving to the U.K. All English participants were taught to write in England.

All 13 features suggested by the experienced examiner as being class characteristics of Polish writing occurred in ≥25% of the Polish sample; however, six of these also occurred in ≥25% of the English samples: the crossed 7, the two-stroke B, and the saw-toothed M, and the A, M, and N starting with an upstroke. Figure 1 illustrates these commonly occurring features and Table 2 specifies the frequencies that these occurred in Polish and English handwriting.

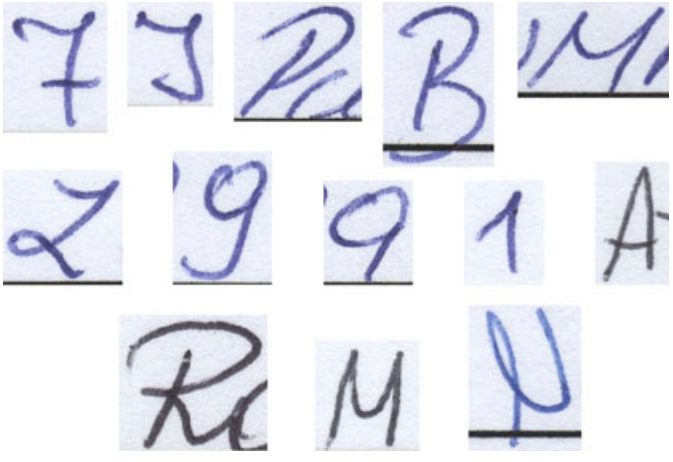


FIG. 1—*Letter and number formations occurring frequently in Polish handwriting.*

Twenty-one characteristics were identified as occurring in ≥3 of the 15 samples examined in detail. Eighteen of the characteristics occurred in ≥25% of the Polish sample as illustrated in Fig. 2 and described in Table 3. One feature, the L with a loop, also occurred in ≥25% of the English samples.

Statistical Analysis of Class Characteristics

In order to discern which of the identified class characteristics occurred significantly more often in the Polish sample compared to the English sample, all of the 34 identified features were analyzed using two-by-two Pearson chi-square tests. Fisher's exact test was used when the expected cell count was below five. The significance cut-off was altered by Bonferroni correction to account for multiple analyses. As 34 features in total were analyzed, the cut-off for statistical significance was therefore adjusted to 0.0015 (0.05/34). Twenty-one characteristics of the 34 analyzed were found to occur significantly more in the Polish samples. Table 4 shows these 21 characteristics, with the results ordered with the features with the highest discriminatory power first, as defined by the chi-square value.

The four most discriminatory features were found to be the b written as a "6" (χ^2 = 51.20, df = 1, p < 0.0005); the d written as a circle and stem (χ^2 = 50.61, df = 1, p < 0.0005); the J with a

TABLE 2—*Description and frequencies of the suggested characteristics.*

Description of Characteristic	Polish N (%)	English N (%)
7 with a cross stroke	35 (87.5%)	24 (60%)
J with loop in corner	34 (85%)	3 (7.5%)
P with 2 strokes	34 (85%)	7 (17.5%)
B with 2 strokes	34 (85%)	21 (52.5%)
M with saw-tooth shape	32 (80%)	25 (62.5%)
Z with loop in bottom corner	31 (77.5%)	6 (15%)
9 with curled bottom	28 (70%)	1 (2.5%)
9 with circle and stem	26 (65%)	0 (0%)
1 with introductory stroke	26 (65%)	1 (2.5%)
A starting with upstroke	25 (62.5%)	16 (40%)
R with 2 strokes	24 (60%)	5 (12.5%)
M starting with upstroke	13 (32.5%)	12 (30%)
N starting with upstroke	12 (30%)	14 (35%)

FIG. 2—*Letter and number formations occurring frequently in Polish handwriting.*

TABLE 3—*Description and frequencies of the characteristics in the exploratory analysis.*

Description of Characteristic	Polish *N* (%)	English *N* (%)
b written as a 6	36 (90%)	4 (10%)
d with circle and stem	31 (77.5%)	0 (0%)
4 written in one stroke	30 (75%)	0 (0%)
p with open bowl	29 (67.5%)	7 (17.5%)
L with loop in corner	29 (67.5%)	16 (40%)
a with circle and stem	24 (60%)	0 (0%)
I written as a J	24 (60%)	0 (0%)
g with circle and stem	23 (57.5%)	1 (2.5%)
q with circle and stem	23 (57.5%)	2 (5%)
z with loop in bottom corner	22 (55%)	1 (2.5%)
n with introductory stroke	21 (52.5%)	1 (2.5%)
m with introductory stroke	17 (42.5%)	0 (0%)
G with long tail	16 (40%)	8 (20%)
w with loop at end	15 (37.5%)	4 (10%)
Y with U and stem	12 (30%)	0 (0%)
Greek "ε"	12 (30%)	4 (10%)
u with introductory stroke	11 (27.5%)	0 (0%)
U with introductory stroke	10 (25%)	0 (0%)
W with introductory stroke	6 (15%)	1 (2.5%)
N with introductory stroke	5 (12.5%)	0 (0%)
M with introductory stroke	5 (12.5%)	4 (10%)

loop ($\chi^2 = 48.32$, df = 1, $p < 0.0005$), and the one-stroke 4 ($\chi^2 = 48.00$, df = 1, $p < 0.0005$). These all occurred in more than 75% of Polish samples and <11% of English samples.

Comparison with Copybook

All features found to occur in ≥25% of the Polish samples and hence defined as class characteristics were compared with the Polish copybook patterns to identify which class characteristics were similar to the copybook and which ones differed significantly from the copybook pattern. In total, 31 features were defined as class characteristics in this way. This included all the features described in Tables 2 and 3 with the exception of the three features occurring <25%: the W, N, and M with introductory strokes. Twenty-one features reflected the copybook patterns and 10 features were departures from the copybook style. Figure 3 illustrates the similar features and Fig. 4 illustrates the class characteristics that did differ from the copybook.

The characteristics that differed contain six features that can be considered "circle and stem" patterns; two that are letters starting with an upstroke; a saw-tooth M; and the one-stroke 4. The Polish samples contained between six and 18 of the 21 copybook derived class characteristics with a median number of 13 of these features (25th–75th percentiles: 11–15).

TABLE 4—*Analysis of features from hypothesized and exploratory exploration.*

Characteristic	Hypothesized/ Exploratory	Polish %	English %	χ^2	*p*
b written as a 6*	E	90	10	51.20	0.000
d with circle and stem*	E	77.5	0	50.61	0.000
J with loop in corner*	H	85	7.5	48.32	0.000
4 written in one stroke*	E	75	0	48.00	0.000
9 with curled bottom*	H	70	2.5	39.43	0.000
9 with circle and stem*	H	65	0	38.52	0.000
P with 2 strokes	H	85	17.5	36.47	0.000
1 with introductory stroke*	H	65	2.5	34.94	0.000
a with circle and stem*	E	60	0	34.29	0.000
I written as a J*	E	60	0	34.29	0.000
Z with loop in bottom corner	H	77.5	15	31.43	0.000
g with circle and stem*	E	57.5	2.5	28.81	0.000
z with loop in bottom*	E	55	2.5	26.91	0.000
q with circle and stem*	E	57.5	5	25.66	0.000
n with introductory stroke*	E	52.5	2.5	25.08	0.000
p with open bowl	E	67.5	17.5	24.44	0.000
m with introductory stroke*	E	42.5	0	21.59	0.000
R with 2 strokes	H	60	12.5	19.53	0.000
Y with U and stem*	E	30	0	14.12	0.000
u with introductory stroke*	E	27.5	0	12.75	0.000
U with introductory stroke*	E	25	0	11.43	0.001

*Included in algorithm.

Development and Testing of Algorithms

The 80 handwriting samples used to identify the class characteristics were used to develop the algorithms with the remaining 26 samples used to test the algorithms. The features that were identified as significantly distinguishing between Polish and English handwriting were ranked according to their chi-square value. Features that were found to occur in ≥11% of the English samples were excluded: the two-stroke P, the Z with a loop in the bottom corner, the p with an open bowl, the two-stroke R. The remaining 17 features were used in the algorithms and are indicated in Table 4.

Four features were found to occur in ≥75% of the Polish writings: the b written as a "6"; the d with a circle and stem; the J with a loop; and the 4 written in one stroke. Nine features occurred in 50–74% of the samples: the 9 with a curled bottom; the 9 with a circle and stem; the 1 with an introductory stroke; the a with a circle and stem; the I written as a J; the g with a circle and stem; the z with a loop; the q with a circle and stem; the n with an introductory stroke. Four features occurred in 25–49% of the samples: the m with an introductory stroke; the Y written as a "u" and a stem; the u with an introductory stroke; the U with an introductory stroke. Scores and cut-offs arising from the three different algorithms as described in Methods were calculated. Table 5 shows the range of scores for the English and Polish samples and the 10th percentile scores from the Polish samples for each algorithm.

The cut-off scores resulting from this analysis were calculated as follows:

FIG. 3—*Class characteristics similar to the copybook pattern.*

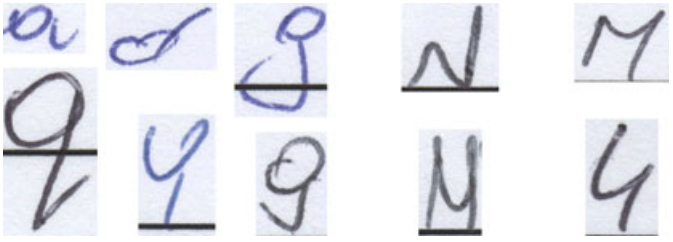


FIG. 4—*Class characteristics differing from the copybook.*

TABLE 5—*Range of scores for the three algorithms.*

	Minimum Score	Maximum Score	10th Percentile
Algorithm 1			
Polish	3	17	5
English	0	2	
Algorithm 2			
Polish	6	21	6
English	0	3	
Algorithm 3			
Polish	9	35	10
English	0	5	

- Algorithm 1: ≤2 not Polish; 3–4 uncertain; ≥5 Polish
- Algorithm 2: ≤3 not Polish; 4–5 uncertain; ≥6 Polish
- Algorithm 3: ≤5 not Polish; 6–9 uncertain; ≥10 Polish

The age ranges, gender, and time spent in the U.K. of the writers of the 13 Polish samples, 12 English samples, and one Slovakian sample are shown in Table 6. The age range for the writers of the Polish samples included younger and older participants than the original sample of 40.

TABLE 6—*Demographics of participants used to test algorithm.*

		Polish (n = 13)	English (n = 12)	Slovakian (n = 1)
Age	Median (25–75‰)	26 years (21.00–32.5)	29 years (22.75–35.00)	35 years
	Range	12–59	21–41	
Gender	Male	6	6	1
	Female	7	6	0
Time in U.K.	Median (25–75‰)	2 years (1–2.25)	n/a	9 years
	Range	0.25–5		
Age moved to U.K.	Median (25–75‰)	23.5 years (20.5–29)		26 years
	Range	11–58		

TABLE 7—*Accuracy of each algorithm for predicting nationality.*

	Not Polish	Uncertain	Polish
Algorithm 1			
Polish	0	1	12
English	12	0	0
Slovakian	0	0	1
Algorithm 2			
Polish	0	0	13
English	12	0	0
Slovakian	0	0	1
Algorithm 3			
Polish	0	0	13
English	12	0	0
Slovakian	0	0	1

The results of the accuracy of the algorithms are shown in Table 7. Algorithms 2 and 3 successfully categorized all of the Polish samples. Algorithm 1 calculated one Polish sample as uncertain. None of the algorithms successfully identified the Slovakian writer as a non-Polish participant.

Discussion

A large number of class characteristics were identified in Polish individuals' handwriting, and class characteristics were identified that were specific to Polish individuals as opposed to English individuals. The influence of the taught writing style was observed in Polish adults' handwritings. An algorithm was developed that was able to successfully distinguish between Polish and English samples.

Identification of Class Characteristics

Using the cut-offs suggested by Berthold and Wooton (19), where ≥25% was considered to be indicative of a characteristic being common or very common in a population, 31 characteristics were identified in the Polish samples that could be regarded as class characteristics. This is comparable to the number of characteristics identified by King (21) in the writings of Mexican immigrants. King (21) uncovered 24 characteristics in the writings of Mexican immigrants to the U.S. using a method similar to the one employed exploratory examination of the current study but using a lower threshold of 10% rather than 25% to accept a feature as a class characteristic.

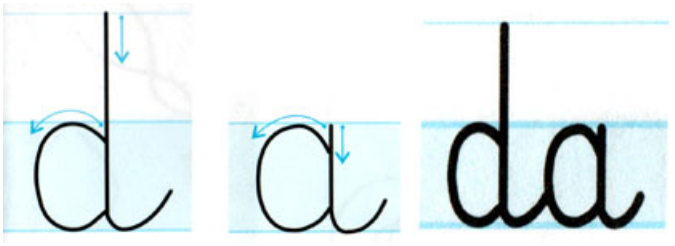


FIG. 5—*Polish copybook instructions for d and a (from [14]).*

Twenty-one features were identified that were common in Polish writing and also occurred significantly more in Polish writing compared to English writing. Of these, 17 characteristics occurred in more than 24% of the Polish samples and less than 11% of the English samples. It is possible that a British document examiner, unfamiliar with Polish handwriting, may consider these characteristics as unusual individual features and, as a result, give them undue weight in a handwriting examination. The characteristics identified in this study can therefore act as a guide to examiners unfamiliar with Polish writing when examining a case where the examiner is aware that the writer may be of Polish origin. For example, features such as the b written as a "6," characters written as a circle and a stem (d, a, 9, g, and q), and I and J written using the same construction, with a loop in the top corner, were all found to be uncommon in English-educated U.K. residents, whereas they were found to be very common in Polish-educated U.K. residents.

It is possible to compare some of the characteristics which were identified in this study with studies that have identified similar characteristics in other groups of writers. For example, the Greek "ε" was found in 30% of the Polish samples and 10% of the English samples. In comparison, this feature has been shown to be rare in U.S.-educated populations (1–3% [15–17]) and common in writers educated in Mexico (26–34% [19,21]).

The approach adopted in this study was to look closely at a subset of 15 samples letter by letter to identify possible characteristics before looking only for these features in the sample as a whole. This method did successfully uncover a large number of distinguishing features of Polish writing; however, a more systematic approach, as taken by King (21), that may have uncovered further class characteristics, would have been to examine the entire Polish sample letter by letter and to have calculated the frequency of each construction of each character. These features could then have been quantified in the English samples and comparisons made as to the relative frequencies of each character construction between the two groups.

Relation of Class Characteristics to Copybook

A strong association was found between the identified class characteristics and the copybook patterns used in Poland. Twenty-one of the 31 characteristics found to occur commonly in Polish writing were found to be of a similar construction to those found in the Polish copybook as illustrated in Fig. 3. Polish samples contained an average of 13 and a minimum of six of these 21 copybook features. This supports the hypothesis that the underlying taught system contributes to the class characteristics of a population of writers (1,8,9,23). Studies that have found weak associations between copybooks and class characteristics have been in populations in which there is a lack of a national copybook, for example in England (10). In contrast, this study has found class characteristics in Polish individuals that do reflect the copybook in many instances. The influence of the copybook in the class characteristics may be because of the apparent use of a single copybook pattern in Poland and an emphasis on penmanship in early education. It is believed from anecdotal evidence of discussing writing skills with Polish-educated adults that penmanship was specifically taught and emphasized as an important skill in the early years of education.

Not all class characteristics, however, reflected the copybook construction. Ten of the features that were identified as distinctive class characteristics in the Polish sample were found to have no relation to the copybook as illustrated in Fig. 4. Departures from the copybook may be influenced by various factors, for example the ease of pen movement in the construction of a character, or the frequent exposure to another style of writing. Further exploration of the patterns of departures from the copybook style is warranted.

One pattern that was observed in this study was that six of the characters that showed a departure from the copybook style shared a common construction: a circle and stem design. Circle and stem patterns have been observed in previous studies in different populations. Livingston (15) described the presence of what he called a "foreign small d," illustrated as a circle and stem, in 0.5% of the handwriting samples in a U.S. population. Circle and stem "9"s have been noted as occurring frequently in the writings of immigrants to the U.S. from various countries (19). One possibility is that this common construction may have arisen from the method in which Polish children are taught to construct and then join letters. Figure 5 illustrates the Polish copybook instructions for constructing a "d" and an "a" and then joining a "d" to an "a." The formations of the "d" and "a" appear to be in two separate strokes. When joining the letters, the instruction is to attach the joining stroke of the first letter to the second letter at the eight o'clock position. This could result in the second letter consisting of a circle starting and ending at eight o'clock followed by a pen-lift and a separate stem as was observed in the circle and stem patterns prevalent in the Polish handwriting samples.

Further influences of the copybook may have been elicited if each copybook construction had been specifically looked for in the samples and frequencies of occurrence of each copybook characteristic in the samples of writing calculated. It would then have been possible to examine the influence of the copybook as a whole and possibly identify patterns in the features from the copybook that were strongly retained or lost in adult handwriting.

Algorithm to Identify Polish Handwriting

The method used to develop the algorithms resulted in one that used 17 class characteristics of Polish writing in order to discriminate between Polish and English handwriting. The discriminatory power of characteristics as defined by the chi-square value was used to rank the features, excluding features that were also fairly common (>10%) in the comparison population. Weight was then added to features that were most prevalent in the samples. This methodology appears to be one that holds promise. Adding weighting to the most discriminatory features aided the accuracy of the algorithms in that the weighted algorithms both managed to successfully identify one Polish sample that the unweighted algorithm had classified as uncertain.

In the test phase, the participants' age range was wider that that of the population used to develop it. The age range of the Polish participants in the test phase was 12–59 years and in the development phase 19–51 years. The participants in the test phase had resided in the U.K. for an average of 2 years, similar to the development phase sample. The ability of the algorithm to accurately identify Polish participants from a wide age range supports its

validity. The apparent consistent use of the same copybook to teach handwriting in Poland has perhaps influenced the homogeneity of Polish writings and permitted the ease of development of the algorithm.

The algorithm was not able to distinguish the handwriting of a Slovakian participant from the Polish participants. This is perhaps not surprising as there are close geographical and cultural connections between Poland and Slovakia and it is possible that there are similar taught writing styles in the two countries. A previous study has commented on the minimal differences that can exist between geographically adjacent countries (24). Further investigation into handwriting from Eastern Europe countries may be able to identify which countries have similar writing styles and those that differ. It may, however, also be possible to engage the same methodology using samples of Polish and Slovakian handwriting in order to identify more subtle differences that may exist between the two countries, perhaps reflecting differences in copybook styles, and create an algorithm that is successful in discriminating between these two countries' handwriting.

Further Work

This study has highlighted various areas that could benefit from further investigation. Further investigation of the prevalence of all the copybook patterns in the writing of Polish adults could clarify further how influential the taught writing style is in the adult handwriting of this population. The identification of patterns of adherence and departure from the copybook constructions may shed light on the factors that influence the maintenance of taught writing styles into adulthood. The way handwriting is taught, and the emphasis on penmanship in schools, may assert an influence on the adherence to the copybook style in adult handwriting. Comparative studies between countries as to the similarities and differences between the taught writing style and adult handwriting, taking a measure of the emphasis on penmanship in schools as an independent variable, could clarify this relationship. The link between the circle and stem patterns and the taught writing system could be examined by comparing taught systems that use a letter linking system similar to the Polish one when teaching cursive writing, and ones that do not, to see if there is a relation between this and the presence of this characteristic in adult handwriting.

It is acknowledged that the amount of writing available may affect the accuracy of the model. In this study, all the writing samples consisted of the same content in which all the upper and lower case letters and numbers were represented. In a sample of writing available to a document examiner in a typical examination, this full range of characters is unlikely to be available. Further testing of the algorithm with more representative samples of writing is necessary to judge whether it is viable for use in casework.

References

1. Simner ML. The origin of class characteristics: an empirical investigation of a major principle in forensic document examination. J Forensic Doc Exam 1998;11:39–50.
2. Found B, Rodgers D. A consideration of the theoretical basis of forensic handwriting examination. Int J Forensic Doc Examiners 1998;4(2):109–18.
3. Huber RA, Headrick AM. Handwriting identification: facts and fundamentals. Boca Raton, FL: CRC Press, 1999.
4. Schuetzner EM. Class characteristics of hand printing. J Am Soc Questioned Doc Examiners 1999;2(1):5–33.
5. Cheng N, Lee GK, Yap BS, Lee LT, Tan SK, Tan KP. Investigation of class characteristics in English handwriting of the three main racial groups: Chinese, Malay and Indian in Singapore. J Forensic Sci 2005;50(1):1–8.
6. Muehlberger RJ. Class characteristics of Hispanic writing in the southeastern United States. J Forensic Sci 1989;34(2):371–6.
7. Hilton O. Scientific examination of questioned documents, Rev. edn. Boca Raton, FL: CRC Press, 1993.
8. Muehlberger RJ, Newman KW, Regent J, Wichmann JG. A statistical examination of selected handwriting characteristics. J Forensic Sci 1977;22(1):206–15.
9. Simner ML, Smits-Englesman BCM. The use of foreign copybook patterns to determine the country of origin of the author of a questioned document. J Forensic Doc Exam 2000;13:45–51.
10. Cha S, Yoon S, Tappert CC. Handwriting copybook style identification for questioned document examination. J Forensic Doc Exam 2005–2006;17:1–14.
11. Hastings S. Handwriting. Times Educ Suppl 2004;12:11.
12. Copybook from Poland, 1974 [computer file]. European Network of Forensic Handwriting Experts Database, 2008, European Network of Forensic Science Institutes, http://www.enfsi.eu/page.php?uid=76.
13. Fajcht H, Miązek D. Uczę się pisać, Zeszyt 1. Gdańska, Poland: Wydawnictwo Educacyjine, Res Polona, 2008.
14. Fajcht H, Miązek D. Uczę się pisać, Zeszyt 2. Gdańska, Poland: Wydawnictwo Educacyjine, Res Polona, 2008.
15. Livingston UB. Frequency of certain characteristics in handwriting pen printing of 200 people. J Forensic Sci 1963;8(2):250–9.
16. Horton RA. A study of the occurrence of certain handwriting characteristics in a random population. Int J Forensic Doc Examiners 1996;2(2):95–102.
17. Zimmerman J. Counting handwriting characteristics. Int J Forensic Doc Examiners 1998;4(4):318–22.
18. Tull P. A study of numbers, their variation and variability: a contemporary study of the variation in the construction of numbers and a comparison with two previous studies. J Forensic Doc Exam 1997;10:41–51.
19. Berthold NN, Wooton EX. Class characteristics of Latin American hand printing. Int J Forensic Doc Examiners 1998;4(2):134–51.
20. Tweedy JS. A study of Hmong handwriting. J Am Soc Questioned Doc Examiners 2000;3(1):11–7.
21. King BH. Class characteristics of Mexican hand printing. J Am Soc Questioned Doc Examiners 2003;6(1):1–33.
22. Strach SJ. Proposed research areas on handwriting comparisons. Int J Forensic Doc Examiners 1998;4(4):310–2.
23. Miller JT. Departure from handwriting system. J Forensic Sci 1972;17(1):107–32.
24. Stanghor GR. Comments on the determination of nationality from handwriting. J Forensic Sci 1971;16(3):343–58.

Additional information and reprint requests:
Dr. Susan J. Turnbull, M.A., D.Clin.Psy., M.Sc.
Forensic Handwriting and Document Investigation
25 Cameron Street
Stonehaven AB39 2BL
U.K.
E-mail: turnbull@fhdi.co.uk

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01454.x
Available online at: interscience.wiley.com

PAPER

QUESTIONED DOCUMENTS; CRIMINALISTICS

Cedric Neumann,[1] *Ph.D. and Pierre Margot,*[2] *Ph.D.*

Considerations on the ASTM Standards 1789-04 and 1422-05 on the Forensic Examination of Ink

ABSTRACT: The ASTM standards on Writing Ink Identification (ASTM 1789-04) and on Writing Ink Comparison (ASTM 1422-05) are the most up-to-date guidelines that have been published on the forensic analysis of ink. The aim of these documents is to cover most aspects of the forensic analysis of ink evidence, from the analysis of ink samples, the comparison of the analytical profile of these samples (with the aim to differentiate them or not), through to the interpretation of the result of the examination of these samples in a forensic context. Significant evolutions in the technology available to forensic scientists, in the quality assurance requirements brought onto them, and in the understanding of frameworks to interpret forensic evidence have been made in recent years. This article reviews the two standards in the light of these evolutions and proposes some practical improvements in terms of the standardization of the analyses, the comparison of ink samples, and the interpretation of ink examination. Some of these suggestions have already been included in a DHS funded project aimed at creating a digital ink library for the United States Secret Service.

KEYWORDS: forensic science, ink examination, standardization, quality assurance, interpretation, ASTM standards

The purpose of the examination of a questioned ink specimen in forensic science is to identify its source (a particular pen, its manufacturer, or its origin), its date (e.g., its date of first production), and how it compares with other inks (questioned or from known source). The inference of the source of an ink specimen can be achieved by the direct comparison of the ink specimen against particular writing instruments (or writings) considered in a case. In some cases, it may be necessary to gather information on the manufacturer of an ink specimen for intelligence purposes. This is achieved by comparing the specimen to an ink library (1–6). These two different examination processes are respectively described as ink comparison and ink identification in two ASTM standards.

The ASTM standards on Writing Ink Identification [ASTM 1789-04 (7)] and on Writing Ink Comparison [ASTM 1422-05 (8)] are the most up-to-date guidelines that have been published on the forensic analysis of ink. These standards are "Standard Guide"; in other words, they are a general outline for the forensic examination of ink and are not binding documents. Nevertheless, they are part of the standard operating procedures (SOPs) of several forensic laboratories performing ink analysis and therefore have a significant impact on the field. Their scope (and they are the only ones to do so) is to cover and offer recommendations on the entire process of the examination of ink in forensic science, namely:

- Analysis: The analysis of selected characteristics of the ink samples;
- Comparison: The comparison of the analyzed characteristics with a view to make a decision on their (lack of) correspondence;
- Evaluation: The forensic interpretation of the (lack of) correspondence between the measured characteristics of two samples.

These standards currently do not reflect the recent evolutions of the forensic field. Also, they do not encapsulate some of the fundamental needs for transparency and objectivity expressed recently (9–13), or the proposed improvements in the interpretation of forensic evidence.

This article explores these standards in light of the theoretical foundations of the identity of source described by several authors (14–17). The following sections consider:

- The selection and the analytical measurement of ink characteristics;
- The comparison and decision-making regarding the correspondence of sets of characteristics acquired from different ink samples;
- The interpretation of the results of these comparisons.

For each section, this report proposes practical solutions and research directions to the objections raised and, thus, improvements to the field of forensic ink analysis. Some of these research directions have only been recently enabled by technological and/or theoretical developments. It is therefore not the aim of this article to lower the merit of past research projects and advancements to the field, but rather to propose a way forward.

The two standards focus heavily on the analysis of ink samples by low- and/or high-resolution thin-layer chromatography (TLC and HPTLC, respectively). Despite the constant application of new analytical techniques to the analysis of ink samples in forensic

[1]The Forensic Science Service, 2920 Trident Court, Birmingham, B377YN, UK.
[2]Ecole des Sciences Criminelles, University of Lausanne, CH-1015 Lausanne, Switzerland.

Received 1 Mar. 2009; and in revised form 8 July 2009; accepted 15 Aug. 2009.

science, TLC and HPTLC are still recognized as the most generally accepted techniques. And as the elements discussed in this article are very similar for these two techniques, the terms HPTLC and TLC will be used indistinctively to represent either or both techniques. Furthermore, based on the state of the art of the literature in ink analysis (18–20), it appears that most of the arguments made in this report will still be valid in the near future for any upcoming analytical technique.

Considerations on the Analytical Process

Based on Smalldon and Moffat (21), Kwan (22) defined some guidelines for the selection of relevant features to characterize evidence items for forensic purpose. As part of these guidelines, Kwan (22) demonstrates that there is a requirement for the constancy and reliability over time of the features measured on evidence items (in our case, ink samples). The level of constancy of these features over a period of interest, or at least the predictability of their evolution, offers the possibility to differentiate between genuine and explainable differences between samples.

Evidently, the constancy of ink characteristics is not only linked to their sensitivity to environmental conditions (e.g., conservation conditions of a document) as demonstrated in (23–29), but is also highly influenced by the quality and reproducibility of the analytical process used to characterize ink samples. It is therefore important to ensure by all means the analytical reproducibility of the ink examination process.

Neither of the aforementioned ASTM standards focuses heavily on this critical issue. In a comparison context, it may be possible to assume and argue, under very restrictive circumstances, that there will be little analytical differences between the simultaneous or quasi-simultaneous HPTLC analyses of questioned specimens and control samples. However, this is certainly not the case in the context of ink identification. Indeed, without a strict quality control of the analytical process, it is impossible to guarantee that the analytical variability between samples analyzed at different times, different locations, and by different operators is minimized. More generally, without quality control, it is not even possible to assert that a given analysis ran as expected and produced meaningful results.

When considering the two ASTM standards, it is noted that the 1798-04 standard on ink identification constantly refers to the 1422-05 standard on ink comparison for matters related to ink analysis. The critical aspect of the quality of the analytical process is not considered thoroughly in the 1789-04 standard, although examiners are rendered attentive in section 7.1.3.3.2 (7) that appropriate materials need to be selected to minimize analytical variability. The 1789-04 standard refers to the 1422-05 standard for all analytical matters. It is then surprising that the only mention on quality control in the 1422-05 standard is the following paragraph:

> 7.7.4.5 Use of a suitable calibration standard is recommended. It should be spotted onto the plate in the same manner. (8)

This paragraph does not indicate what the calibration standard should be, or how it could be used to minimize analytical variability.

Other elements, which are specific to the quality assurance of the considered analytical technique (i.e., TLC or HPTLC), may also be noted in the 1422-05 standard. For example, the use of the word "*approximately*" is certainly a source of concern in a standard that aims at comparing subtle differences between forensic grade samples. More specifically, the amount of material to be extracted, the extraction procedure, the deposition of samples on the silica plates, their elution distance and time, and the choice of the hardware equipment are all left to the individual appreciation of each examiner. The standard implies that the variability of these parameters does not affect the outcome of the analysis, and by extension of the case at hand, and that this variability is therefore acceptable. However, this assumption needs to be validated and/or, if it exists, the outcome of such study needs to be published.

The variability introduced by these parameters in the ink analysis process is both obvious and easy to address to be reduced. It is therefore expected that such elements are specified in the SOPs that are based on these standards. Nevertheless, literature searches on the subject have demonstrated the lack of research and publication in this specific area. No study has been undertaken on detection limits or analytical reproducibility of the forensic analysis of ink samples for identification and comparison purposes. It could be possible to consider the transferability of similar results from other studies performed on the quantitative analysis of other colored material. However, this remains to be done and validated.

As the entire examination process relies on the analytical stage, it is therefore important for the ink examiner community to investigate and propose solutions to measure, control, and minimize the variability between repetitive analyses of ink samples, as suggested in (30).

Considerations on the Match Versus Nonmatch Decision-making Process

Both standards have centered the outcome of the comparison process on the decision of whether questioned specimens and reference samples "match."

The concept of a match is convenient, as it obliges the examiner to take a binary decision on the qualitative identity of two samples, and thus eases the interpretation of the outcome of the comparison in a forensic context. An examiner can declare that two samples are qualitatively identical after having reviewed, explained, and thus dismissed discordances. During this process, examiners in essence force the constancy of the features, described in the previous section, and can then focus on the determination of their evidential value during the next phase.

The concept of match is associated with the "fall off the cliff" effect (31): either the samples match or do not match. This effect, which is linked to all binary decision-making processes, opens the possibility of having false associations and false exclusions between samples. It is therefore necessary to control as much as possible the root cause of these errors, which is, in this case, the capability of ink examiners to review and differentiate between genuine and explainable differences between ink samples.

In both standards, the information needed by ink practitioners to inform decisions on the analytical similarity/dissimilarity (i.e., match vs. nonmatch) between ink samples is spread over three different sections (sections 5, 7, and 9). This can be inconvenient and confusing.

Section 5 of the 1422-05 standard lists a large variety of possible causes to explain differences. While this section lists the causes of possible variations, it does not provide guidance on how to distinguish between (i) those differences that derive from either of the several causes listed and (ii) those differences that result from the inks being from two different sources, with genuinely different characteristics. Section 7 of the 1422-05 standard does not inform

further on how to distinguish between explicable and genuine differences. Indeed, it only expands on the cause of the explicable differences and does not describe their effect. Section 9 is not more helpful as it overlaps with section 7. These sections are reviewed in more details below.

Explicable Differences

Regarding explicable differences, the 1789-04 standard mentions that:

> 7.2.5.4.1 [...] Explicable differences include characteristics arising from diffusion of fluorescent components, differences in the paper controls, differences in color due to fading either of the inks or of the components on the TLC sheet/plate, solvent depletion, or a combination of these and other factors. (7)

An extensive list of these differences can be found in section 5 of the 1422-05 standard. At the beginning of this section, it is stated that:

> 5.1 Most interferences with ink examinations come from variables that interact with the ink. These can be part of the writing process [...], or various forms of contamination on the document, [...]. Simple precautions can usually avoid problems. (8)

A list of general points that examiners should consider to explain differences at the time of interpreting the results of ink analyses follows paragraph 5.1. A large amount of these points relate to events that are not directly linked to the analyses *per se*, but to the past history of the questioned documents. It is noteworthy that, according to the standard, forensic scientists could avoid problems by taking "*simple precautions.*" These precautions aim to ease the interpretation of explainable differences created by complex parameters, such as:

- Modifications of the inks caused by elements that are out of the control of examiners, such as the conservation conditions of questioned documents.
- Modifications that originate from an undetermined past and certainly not at the time of the analysis;
- Modifications, which have barely been quantified or even qualified scientifically.

In other words, these modifications are very complex ones and can have varying levels of severity. Both standards focus on the cause of these modifications and do not describe their consequences. Indeed, the impact of these parameters on the ink sample is very difficult to evaluate a posteriori without data originating from fundamental research. In the absence of the necessary research [with the exception of a few precursors (25–29,32)], it is therefore currently extremely difficult to take these parameters and their impact into account in the match decision-making.

This has been confirmed by routine examination of ink evidence, which has shown that the handling and the interpretation of ink comparisons, with the aim of associating/dissociating samples based on their chemical profile, represent a more complex challenge than presented in the standards. A deeper and more formal scientific approach is needed to investigate the impact of these variables. Indeed, this important topic cannot simply be dealt with using "*simple precautions*," and past studies need to be expanded (25–29,32).

Genuine Differences

Concerning genuine differences, the 1422-05 standards reports that:

> 7.7.7.1 Sample of ink with qualitatively different colorant compositions can be easily distinguished by comparison of the characteristics observed in 7.7.6. (8) (i.e., coloration, migration distance, and relative concentration of the ink components).

And that:

> 9.2.1 If significant, reproducible, inexplicable differences between ink samples are found [...], it may be concluded that the inks do not have a common origin. (8)

Firstly, the location of the elements present in section 9.2.1 is surprising because section 9 addresses issues with respect to the interpretation of the "match/nonmatch" result of ink comparisons in the forensic context, as opposed to section 7.7.7, which considers how to reach that result. In its current form, section 9.2.1 addresses elements that can be considered to be part of the decision-making of whether the chemical profiles of two samples match or not, as well as the implication of this difference in the forensic context. This can be confusing.

Secondly, mirroring the sections of the 1422-05 standard quoted previously on explicable differences, the sections on genuine differences are not more helpful. They do not provide any information on what would be a "*significant*" and "*inexplicable*" difference (e.g., between the relative concentrations of ink components in two different samples). In addition, the use of the term "*reproducible*" is confusing:

- First, an explicable difference may perfectly well be reproducible.
- Secondly, if a difference is not reproducible, does it rather not make it even less explicable? Would an important difference then be dismissed on the grounds that it is not reproducible in 1 of 10 repetitive analyses?

Overall, none of the terms "*significant*," "*reproducible*," and "*inexplicable*" can be quantified or qualified based on the reading of the standard or on the bases of the existing literature in the field.

Concluding on the subject of the comparison of ink samples, neither standard allows examiners to draw clear quantitative and/or qualitative guidelines to differentiate between genuine and explicable differences. Examiners need to rely on their training and experience to take decisions on the match/nonmatch of pairs of samples. Given the binary nature of the match/nonmatch decision, this can only lead to different examiners making opposing decisions for same pairs of samples and thus, to errors.

The authors do not claim by any means that the error rate of ink examination is of concern. However, in the light of recent requirements for greater transparency, better quality assurance (13), and to facilitate the training of new examiners, structured and systematic research on the effect of various environmental and analytical parameters on ink samples, as well as on ways of ensuring a greater objectivity during the comparison process, need to be undertaken and supported by the profession.

Considerations on the Interpretation of the Results of the Analyses

The aim of forensic ink examination is generally to answer questions of interest to legal disputes. A straightforward interpretation

of this goal requires that ink examiners should answer questions of interest with relevant and useful answers. From this perspective, the reading of the two standards shows that this is currently a challenge in the field of the forensic analysis of ink.

In the comparison and identification processes, the field of forensic ink analysis is defined as being focused on the inference of the identity of the source of questioned ink specimens. The scope of both standards reflects this fundamental problematic:

> 1.1 This guide is intended to assist forensic examiners comparing writing or marking inks. Included in this analysis scheme are the necessary tools and techniques available to reach conclusions as to the common or different origin of two samples of ink. (7,8)

Nevertheless, both standards have difficulties to propose informative answers to the source question. The reasons for that are similar in both standards:

- The lack of formalization of the definition of the source of an ink in the two different contexts;
- The absence of a logical framework for the interpretation of ink examinations in both forensic contexts, which would encapsulate the selected definitions of the source of an ink.

These two elements are considered below.

Definition of the Source of an Ink

The 1422-05 standard first defines the source of an ink as the same formula, or "*the same writing instrument or ink well*" (paragraph 4.1). However, paragraph 4.3 then specifies that:

> 4.3 […] When dealing with contemporary inks, however, a match of ink samples […] would not be sufficient to support a definite opinion of common origin. Contemporary ink rarely has sufficient individuality to support a determination of common origin at less than the manufacturing batch level. (8)

Taking it from there, paragraphs 4.1 and 4.3 (and hence the 1422-05 standard) seem to define the source of an ink sample as either one of the following:

(a) An ink manufacturer.
(b) An ink formula.
(c) An ink batch.
(d) A writing instrument.

It is clear that "*contemporary ink rarely has sufficient individuality*" to reach a conclusion of numerical identity of instrument based on the sole analytical process (paragraph 4.3). However, Kwan (22) has shown that the identification process can be constituted of an analytical process, which can be supplemented by a statistical process. It would therefore be incorrect to rule out *a priori* the possibility, in the context of a forensic case, that the source of an ink entry can be defined as a given instrument. This possibility needs to be formally taken into account, and its relevance assessed against the other definitions of the source. Indeed, it will be demonstrated below that it is an extremely important one.

In any case, section 4.3 is not clear as to which definition is actually considered and attributed to the source of an ink. And if all definitions are considered in turn (i.e., depending on the situation), it fails to explicit (or refer to an explicit description) under which circumstances a particular meaning needs to be used over the other ones. This paragraph needs to be more formal and transparent and orient ink examiners when addressing core interpretation issues. The indiscriminate use of multiple definitions of the meaning attributed to the source of an ink, and the lack of formalism throughout both standards render incoherent the section treating the reporting conclusions (section 9) in both standards.

The Interpretative Framework when the Ink Samples do not Match

In preamble, while section 9 should focus on the interpretation of the results of the ink examination process in the forensic context, it appears that section 9 also addresses elements around the decision of matching, or not, samples based on the results of the analytical process. Those elements unnecessarily overlap the ones in section 7.7.7, and confuse the message of section 9. This was discussed above for section 9.2.1. Section 9.2.2, which describes how to report the differentiation between ink samples, is perplexing for the same reason.

> 9.2.2 However, when inks give differing test results, the possibility of batch-to-batch variation within an ink formula must be considered […]. (8)

This statement is correct from the point of view of the interpretation of the comparison between two samples and the decision-making on their match based on their analytical profile: two samples with minor differences may well be of different batches from the same formula. However, this statement should then be present in section 7.7.7, which deals with this issue. This would allow section 9 to focus on the reporting issues.

Furthermore, when it comes to the interpretation of the meaning of an ink comparison in the context of a case, the statement made in section 9.2.2 is irrelevant in most cases:

- Assuming that the source of an ink is considered to be a particular instrument or a particular batch, as described in sections 4.1 and 4.3, if batch-to-batch variations are observed between a questioned specimen and a control sample, then the nonidentity of source is proven and paragraph 9.2.2 is uninformative.
- Assuming the source of an ink is considered to be an ink manufacturer, batch-to-batch variations exist within batches of a specific manufacturer and between batches of different manufacturers; the simple observation of these variations does not assist to infer the identity of the ink manufacturer, as opposed to the knowledge on how to differentiate between these two types of variations. Such knowledge would need to be based on rigorous scientific studies.
- Similarly, assuming that the source of an ink is a particular formula, one may inquire about the extent of difference needed between two batches to consider them as being from different formulation. As it is not possible to answer this question without systematic studies, the question of batch-to-batch variations may not be relevant at all in a forensic context after all.

In conclusion,

- should the definition of an ink be properly stated earlier in either standard,
- should examiners be supported during the selection of the relevant definition within the context of their cases,
- and should elements, addressing the decision of whether the chemical profiles of ink samples are similar or not, be located in the section addressing this issue (section 7.7.7),

then, the entire section 9.2 becomes irrelevant. Indeed, irrespective of the definition of the source of an ink, dissimilar (nonmatching) ink samples cannot originate from the same source within the context of a case.

The Interpretative Framework when the Samples Match

The same comments made for section 9.2 of the standard 1422-05, with respect to the overlapping with section 7.7.7 of this standard, can be made for section 9.3. Furthermore, section 9.3.3 seems to contradict the scope of the standard itself:

> 9.3.3 Reports of conclusions should never state that two ink samples are identical or the same ink. Statements must be within the limits of 9.3.1. (8)

Paragraph 9.3.3 is mistaken in stating that two inks should never be reported as identical. Indeed, two ink samples can be perfectly identical in terms of ink formulation and different in terms of writing instrument. The answer to the scope of the standard depends on the definition of the source.

While there are often not enough characteristics to identify an instrument or a batch solely based on the analysis of an ink sample, the inference of the identity of source can be supported by a statistical process (22). This statistical process would supplement the limitations in the discriminating power of the chosen analytical techniques. In such case, the use of relevant databases, populated depending on the chosen definition of the source of an ink, can inform the interpretation of the result of ink examinations in forensic contexts. While this interpretative process would not lead to the individualization of a given instrument, probabilistic statements could be reported for ink evidence as they are for other evidence types.

At this point, it is important to realize that some of the definitions of the meaning attributed to the source of an ink can make the examination process uninformative from the forensic point of view. Indeed, the fact that a questioned specimen is of the same formula as a control sample does not provide any indication on the actual evidential value of the results of the performed comparison.

We recognize that the determination that two ink samples are not of the same formula can provide valuable information to the judicial system. However, not being able to provide extended information where there is actually a correspondence limits the potential of ink evidence and certainly confuses the actors of the legal system, who are accustomed to gaining such information with other evidence types. Without knowledge on the number of instruments carrying an ink with that formula or on the availability of such an instrument at the time of the creation of the questioned document, the results of the comparison bear relatively limited value to the judicial system.

Paragraph 4.7 proposes to use an ink library to assess the "*rarity of a formula.*" This is certainly a first step toward the interpretation of ink evidence within the forensic context; however, it raises two questions:

- Does a "rare formula" imply that ink of this formula is not found in many instruments, regardless on how many manufacturers produced it?
- Does a "rare formula" imply that ink of this formula is only produced by a handful of manufacturers, regardless of the quantity that is actually produced?

The answers to these questions are very different, and only the first one is relevant to address the source questions as laid out in paragraph 4.1.

The correct interpretation of the outcome of ink comparisons requires the consistent definition and use of the meaning attributed to the source of the ink throughout the entire comparison process, even in the creation of the relevant ink libraries. This aspect is left out from the 1422-05 standard and has not been dealt with in the relevant literature. The development of a formal framework for the interpretation of ink comparison is needed. Indeed, answering positively the first question presented above clearly implies that the source of an ink in the comparison context should be defined as an instrument: ink examiners when interpreting the value of ink evidence in the context of a case are concerned with the determination of the frequency of a particular ink. This frequency can be easily inferred from the proportion of instruments containing that ink, and available at the time of the creation of the document. Creating statistically representative databases of writing instruments can provide the required frequency data. Nevertheless, this requires for the community to accept a departure from defining the source of an ink as a formula, a batch, or a manufacturer in the context of the interpretation of the forensic value of ink comparisons.

Population of Reference Databases

The lack of formalism in the definition of the meaning of the source of an ink in the identification context similarly impacts the 1789-04 standard. The source of an ink is never truly defined properly and the standard sways between (i) considering the source as a formula and (ii) considering the source as a manufacturer. Potential guidelines regarding the population of the ink libraries are affected by this lack of formalism. The 1789-04 considers in turn that the identification process aims at discriminating ink manufacturers (e.g., paragraph 4.1) or ink formulas (e.g., paragraph 4.1.1 or 7.1.1.1). Similarly, section 7, which provides guidance on the constitution of ink libraries, proposes to reference together ink obtained from manufacturers and ink obtained in the retail market. This does not appear to be very coherent considering the purpose of the standard:

- Assuming that the source of an ink is a formula, it is not necessary to have the same formula several times from different manufacturers. When the formulas are not identical *per se*, but are indistinguishable by the analytical technique currently used to construct the collection, it may be necessary to have control samples from each of the formulas. These samples need to be in raw format to allow for their future analysis by a more discriminative technique. It is also necessary to mention that ink formula cannot readily and/or easily be obtained from pens sampled on the retail market because pen manufacturers may not disclose the required information.
- Assuming the source of an ink is an ink manufacturer, the collection of instruments from the retail market assumes one-to-one correspondence between ink and instrument manufacturers. This is not the case (30). In this context, the full cooperation of the ink/instrument manufacturers is also required. While this may be achievable in the case of local manufacturers, it is most often not possible for remote ones. In the evolving and global ink market, this is a clear obstacle to the constitution of a relevant ink collection when the source of an ink is defined as an ink manufacturer. The forensic laboratory from the United States Secret Service is the only entity that seeks the cooperation from worldwide ink manufacturers and had some success in this task. However, these difficulties will mostly prevent the creation of ink libraries in other agencies.

Overall, these definitions do not appear to be very relevant to the forensic problem, nor does it appear that forensic examiners can possibly and practically rely on them to structure most of their investigations.

Considering the source of an ink as a formula may be relevant in an ink dating context because it allows the determination of the date of the first introduction of a particular formula; however, it does not appear very practical in the identification or comparison context. Defining the source of an ink as an ink manufacturer does not provide more information than if it was defined as a formula, and it has been shown that it is no more practical from the point of view of the practitioners. The date of first introduction may be different for each specific manufacturer; nevertheless this information is not helpful in an ink dating context because only the earliest date is relevant. As there are no univocal relationships between ink manufactures and instrument manufacturers, the determination of the ink manufacturer has little investigative value. This is unnecessarily limiting the potential of ink evidence as an investigative tool.

We advocate that the collection of instruments from local retail markets is the only efficient, practical, and friendly way for forensic ink examiners to constitute relevant ink collections and to provide useful investigative information in the context of ink identification (7). However, this strategy implies that it is necessary to consider the source of an ink as a brand/model of an instrument. This is not incompatible with current practice and would only require minor changes.

Conclusions on the Review of the Two Standards and Way Forward

The ASTM 1422-05 and 1789-04 standards (7,8) focus on most aspects of the forensic examination of ink, from the analytical process to the inference of identity of source of ink samples. The focus on these standards has been brought by the fact that they are the only published documents rightly attempting to propose a global answer to ink analysis and interpretation. Nevertheless, their authors have not been helped by the lack of theoretical framework and the lack of structured research in the field of ink analysis. The review of these standards reflects that the ink analytical process can be improved from a quality point of view and that the current framework for the interpretation of the results of ink analyses needs to be completed.

The standardization and calibration possibilities of the proposed analytical techniques need to be considered. The exchange and comparison of ink data analyzed at different times and places and by multiple examiners need to be guaranteed. It is especially important for TLC and HPTLC, where methodologies developed in the examination of other materials (33–35), or developed for DNA and successfully applied to inks (30,32), could be used. Nevertheless, standardization is a wider issue and should also be considered for all other techniques routinely applied or proposed to the examination of ink evidence.

Research in the field of forensic ink examination should not only focus on the application of new analytical technique and on their discrimination capacities on ink sample. The benefits and/or improvements of the proposed research should be expressed clearly in terms of the overall ink examination process. More importantly, the research projects need to consider issues such as detection limits, calibration, standardization, and variability between several analyses of a same ink sample. This is critical to bring ink examination to the standard currently required from the forensic profession (13).

The decision-making process regarding the qualitative identity of ink samples needs to be made transparent, as recently required by US Courts, legal scholars, and the US Academy of Sciences (11–13). The differentiation between explainable and genuine differences needs to be supported by empirical data from the latest publications on the subject (23–29). The answer to the question of "how much difference is enough?" (assuming that a binary decision is what is sought in the first place) can be answered using well-understood statistical models and tests (36), providing that the relevant data is acquired (32). The use of improved guidelines will then prevent multiple examiners from contradicting each other regarding the qualitative identity of pairs of samples, hence avoiding reporting confusing and misleading evidence.

Core interpretation issues need to be more formally addressed in both standards. The inclusion of a structured interpretative framework, as advocated for other evidence types (37–46), will support ink examiners when answering questions as expert witnesses and building relevant ink collection, in both the comparison and identification contexts. The development and the application of this framework will show that traditional explanations relative to the source of an ink, such as a particular batch, a specified ink manufacturer, or an ink formula, have little relevance in most cases.

The definition of the source of an ink as a "brand" or "model" in the identification context, and as "writing instrument" in the comparison context, will allow ink examiners to provide informative and relevant answers, as opposed to the current use of "ink manufacturer," "ink formula," or "ink batch" (47).

The constraints set by the logical framework on the constitution of ink reference collection will help ink examiners to structure the sampling of the ink retail market and the population of reference ink libraries in the identification context. We appreciate that the definition of the source of an ink as a given instrument represents a significant change for practitioners dealing with ink evidence in a comparison context. However, using this definition enables the quantification of the weight of ink evidence (47), while the current situation only considers the "nondifferentiation" of ink samples as the best possible answer to judicial queries.

These standards were last updated several years ago, at a time when very few people in the field were considering some of the aspects presented in this report. Recent research (30,32,47) has aimed at providing concrete solutions to these challenges and could to be used to support the rewriting of these standards to maximize the benefits from the use of ink evidence in forensic science. In fact, taking some precedence on an eventual rewriting of the standards, the recent implementation of the digital ink library at the United States Secret Service (48) has demonstrated the feasibility of the implementation of the proposed concepts (30,32,47) into practice.

References

1. Ellen DM. The scientific examination of documents, 2nd edn. London: Taylor & Francis, 1997.
2. Brunelle RL, Reed RW. Forensic examination of ink and paper. Springfield, IL: C.C. Thomas, 1984.
3. Breedlove CH. The analysis of ball-point inks for forensic purposes. J Chem Educ 1989;66:170–1.
4. Brunelle RL. Ink analysis. In: Siegel J, Knupfer G, Saukko P, editors. Encyclopedia of forensic sciences. San Diego, CA: Academic Press, 2000;591–7.
5. Neumann C, Mazzella WD. Questioned document. In: Worsfold P, Townshend A, Poole C, editors. Encyclopedia of analytical science. London, Oxford, Boston, New York and San Diego: Academic Press, 2004;465–71.

6. ASTM International. ASTM E444-07 standard guide for scope of work of forensic document examiners. West Conshohocken, PA: ASTM International, 2007.
7. ASTM International. ASTM E1789-04 standard guide for writing ink identification. West Conshohocken, PA: ASTM International, 2004.
8. ASTM International. ASTM E1422-05 standard guide for test methods for forensic writing ink comparison. West Conshohocken, PA: ASTM International, 2005.
9. Saks MJ, Koehler JJ. The individualization fallacy in forensic science evidence. Vand L Rev 2008;61:199–219.
10. Saks MJ. Merlin and Solomon: lessons from the law's formative encounters with forensic identification science. Hastings Law J 1998;49:1069–141.
11. Saks MJ, Koehler JJ. The coming paradigm shift in forensic identification science. Science 2005;309:892–5.
12. Saks MJ. The legal and scientific evaluation of forensic science (especially fingerprint expert testimony). Seton Hall Law Rev 2003;33:1167–87.
13. Committee on Identifying the Needs of the Forensic Sciences Community, National Research Council. Strengthening forensic science in the United States: a path forward. Washington, DC: National Academies Press, 2009.
14. Lindley DV. A problem in forensic science. Biometrika 1977;64:207–13.
15. Evett IW. A Bayesian approach to the problem of interpreting glass evidence in forensic science casework. J Forensic Sci Soc 1986;26:3–18.
16. Robertson B, Vignaux GA. Interpreting evidence: evaluating forensic science in the courtroom. Chichester, England: John Wiley and Sons, 1995.
17. Champod C. Identification/individualisation: overview and meaning of ID. In: Siegel J, Knupfer G, Saukko P, editors. Encyclopedia of forensic sciences. San Diego, CA: Academic Press, 2000;1077–83.
18. Tebbett IR. Chromatographic analysis of inks for forensic science applications. Forensic Sci Rev 1991;3:71–82.
19. Zlotnick JA, Smith FP. Chromatographic and electrophoretic approaches in ink analysis [review]. J Chromatogr B Biomed Sci Appl 1999;733:265–72.
20. Neumann C, Margot P. Forensic ink analysis: a review of the techniques. Rev Int Criminol Police Tech 2003;56:341–60.
21. Smalldon KW, Moffat AC. Calculation of discriminating power for a series of correlated attributes. J Forensic Sci Soc 1973;13:291–5.
22. Kwan QY. Inference of identity of source [PhD thesis]. Berkeley, CA: Department of Forensic Science, University of California, 1977.
23. Weyermann C, Kirsh D, Costa-Vera C, Spengler B. A GC/MS study of the drying of ballpoint pen ink on paper. Forensic Sci Int 2006;168:119–27.
24. Brazeau L, Gaudreau M. Ballpoint pen inks: the quantitative analysis of ink solvents on paper by solid-phase microextraction. J Forensic Sci 2007;52:209–15.
25. Hamed HR, Safey El-Din NM, El-Laithy SA, Mansour OY, Sabaa MW. Effect of accelerated fading on the stability of inks marked on different types of papers. International Journal of Forensic Document Examiners 1997;3:229–36.
26. Andrasko J. HPLC analysis of ballpoint pen inks stored at different light conditions. J Forensic Sci 2001;46:21–30.
27. Andrasko J. Changes in composition of ballpoint pen inks on aging in darkness. J Forensic Sci 2002;47:324–7.
28. Andrasko J, Kunicki M. Inhomogeneity and aging of ballpoint pen inks inside of pen cartridges. J Forensic Sci 2005;50:542–7.
29. Weyermann W, Kirsh D, Costa-Vera D, Spengler B. Photofading of ballpoint dyes studied on paper by LDI and MALDI MS. J Am Soc Mass Spectrom 2006;17:297–306.
30. Neumann C, Margot P. New perspectives in the use of ink evidence in forensic science - part I: development of a quality assurance process for forensic ink analysis by HPTLC. Forensic Sci Int 2009;185:29–37.
31. Finkelstein MO, Fairley WB. A Bayesian approach to identification evidence. Harv L Rev 1970;83:489–517.
32. Neumann C, Margot P. New perspectives in the use of ink evidence in forensic science - Part II: development and testing of mathematical algorithms for the automatic comparison of ink samples analysed by HPTLC. Forensic Sci Int 2009;185:38–50.
33. Stead AH, Gill R, Wright T, Gibbs JP, Moffat AC. Standardised thin-layer chromatographic systems for the identification of drugs and poisons. Analyst 1982;107:1106–68.
34. Ambrus A, Fuezesi I, Lantos J, Korsos I, Szathmary M, Hatfaludi T. Application of TLC for confirmation and screening of pesticide residues in fruits, vegetables, and cereal grains: part 2: repeatability and reproducibility of Rf and MDQ values. J Environ Sci Health B 2005;40:485–511.
35. Gill R, Law B, Brown C, Moffat AC. A Computer search system for the identification of drugs using a combination of thin-layer chromatographic, gas-liquid chromatographic and ultraviolet spectroscopic data. Analyst 1985;110:1059–65.
36. Mansfield AJ, Wayman JL. Best practices in testing and reporting performance of biometric devices v2.0.1, NPL report CMSC 14/02. Nottinghalm, UK: Center for Mathematics and Scientific Computing, UK National Physical Laboratory, 2002.
37. Neumann C, Champod C, Puch-Solis R, Egli N, Anthonioz A, Bromage-Griffiths A. Computation of likelihood ratios in fingerprint identification for configurations of any number of minutiae. J Forensic Sci 2007;2007:54–64.
38. Gonzalez-Rodriguez J, Drygajlo A, Ramos-Castro D, Garcia-Gomar M, Ortega-Garcia J. Bayesian analysis of fingerprint, face and signature evidence with automatic biometrics systems. Forensic Sci Int 2005;155:126–40.
39. Gonzalez-Rodriguez J, Drygajlo A, Ramos-Castro D, Garcia-Gomar M, Ortega-Garcia J. Robust estimation, interpretation and assessment of likelihood ratios in forensic speaker recognition. Computer Speech Lang 2006;20:331–55.
40. Evett IW, Weir BS. Interpreting DNA evidence. Sunderland, MA: Sinauer, 1998.
41. Roux C, Margot P. An attempt to assess the relevance of textile fibres recovered from car seats. Sci Justice 1997;37:225–30.
42. McDermott SD, Willis SM, McCullough JP. The evidential value of paint. Part II: a Bayesian approach. J Forensic Sci 1999;44:263–9.
43. Curran JM, Hicks-Champod TN, Buckleton JS. Forensic interpretation of glass evidence. Boca Raton, FL: CRC Press, 2000.
44. Evett IW, Champod C. A probabilistic approach to fingerprint evidence. J Forensic Ident 2001;51:101–22.
45. Champod C, Evett IW, Kuchler B. Earmarks as evidence: a critical review. J Forensic Sci 2001;46:1275–84.
46. Champod C, Baldwin D, Taroni F, Buckleton JS. Firearms and tool marks identification: the Bayesian approach. The AFTE J 2003;35:307–16.
47. Neumann C. New perspectives in the use of ink evidence in forensic science [PhD thesis]. Lausanne, Switzerland: University of Lausanne, 2008.
48. Neumann C, CAMAG Inc, Champod C. The International Ink Library of the United States Secret Service: a new and efficient way of managing the data, DHSARPA, Contract HSHQDC-06-R-00066, 2006.

Additional information and reprint requests:
Cedric Neumann, Ph.D.
Ecole des Sciences Criminelles
University of Lausanne
CH-1015 Lausanne
Switzerland
E-mail: cedric.neumann@me.com

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01457.x
Available online at: interscience.wiley.com

PAPER

TOXICOLOGY

Ushtana Antia,[1] *M.Sc.; Malcolm D. Tingle,*[2] *Ph.D.; and Bruce R. Russell,*[1] *Ph.D.*

Validation of an LC–MS Method for the Detection and Quantification of BZP and TFMPP and their Hydroxylated Metabolites in Human Plasma and its Application to the Pharmacokinetic Study of TFMPP in Humans*

ABSTRACT: An LC–MS method was developed for benzylpiperazine (BZP) and trifluoromethylphenylpiperazine (TFMPP), constituents of "party pills" or "legal herbal highs," and their metabolites in human blood plasma. Compounds were resolved using a mixture of ammonium formate (pH 4.5, 0.01 M) and acetonitrile (flow rate of 1.0 mL/min) with a C18 column. Calibration curves were linear from 1 to 50 ng/mL ($R^2 > 0.99$); the lower limit of quantification (LLOQ) was 5 ng/mL; the accuracy was >90%; the intra- and interday relative standard deviations (R.S.D) were <5% and <10%, respectively. Human plasma concentrations of TFMPP were measured in blood samples taken from healthy adults ($n = 6$) over 24 h following a 60-mg oral dose of TFMPP: these peaked at 24.10 ng/mL (±1.8 ng/mL) (C_{max}) after 90 min (T_{max}). Plasma concentrations of 1-(3-trifluoromethyl-4-hydroxyphenyl) piperazine peaked at 20.2 ng/mL (±4.6 ng/mL) after 90 min. TFMPP had two disposition phases ($t_{1/2}$ = 2.04 h (±0.19 h) and 5.95 h (±1.63 h). Apparent clearance (Cl/F) was 384 L/h (±45 L/h).

KEYWORDS: forensic science, toxicology, liquid chromatography–mass spectrometry, BZP, TFMPP, pharmacokinetics, metabolism

Piperazine-based compounds such as trifluoromethylphenylpiperazine (TFMPP), benzylpiperazine (BZP), methoxyphenylpiperazine (MeOPP), chlorophenylpiperazine (mCPP), and fluorophenylpiperazine (pFPP) are used extensively as recreational drugs around the world. BZP and TFMPP are the most commonly encountered active constituents of "party pills" or "herbal highs." Party pills are frequently sold over the internet and often have varying types and amounts of the active ingredients, with names like "Charge" (50 mg BZP and 200 mg TFMPP), "Bliss" (100 mg BZP and 50 mg TFMPP), or "Mash" (37.5 mg pFPP) to target potential consumers.

There is little information available about the pharmacological effects and metabolism of BZP and TFMPP. Unlike many other psychostimulant drugs such as the amphetamines that were developed as medications with extensive testing prior to their use in humans, the effects and metabolism of piperazine-based drugs were relatively unknown prior to their rise in popularity as recreational drugs.

Since the late 1990s and the popularity of piperazine derivatives as recreational drugs, a number of studies have sought to fill the gap in knowledge surrounding these compounds. Currently, studies have implicated major cytochrome P450 enzymes in the metabolism of BZP and TFMPP (1–3), and their neural effects in humans have also been studied using electroencephalography (4–6). The effects of piperazine analogs on the metabolism of commonly used medications have also been investigated (3,7).

To conduct a study on the pharmacokinetics of BZP and TFMPP in humans, a sensitive and selective assay utilizing high-performance liquid chromatography coupled with mass spectrometry (LC–MS) was developed for the detection and quantification of the BZP and TFMPP in human blood plasma.

This method was developed in accordance with the industry guidelines for the validation of bioanalytical methods (8). The validation of the chemical assay included determination of fundamental parameters of accuracy, precision, selectivity, sensitivity, reproducibility, and stability. The validation process was complicated by the simultaneous separation and detection of five closely related compounds. To ensure the validity of this method for pharmacokinetic samples, the validation process was carried out for plasma samples spiked with a mixture containing all five analytes. To ensure the identity of each peak and the selectivity of the method, individual spikes of each compound were also used. As a large number of pharmacokinetic samples were to be analyzed by this method, a short run time was required.

This method has been used to analyze samples collected from a human pharmacokinetic trial. The pharmacokinetic parameters of TFMPP are described below. TFMPP doses can range from 5 mg to over 100 mg. A dose of 60 mg was chosen for this study as it

[1]School of Pharmacy, University of Auckland, Private Bag 92019, Auckland, New Zealand.
[2]Department of Pharmacology, University of Auckland, Private Bag 92019, Auckland, New Zealand.
*Funding for this research was received from the Health Research Council of New Zealand.

Received 5 April 2009; and in revised form 1 July 2009; accepted 5 Aug. 2009.

is a common dosage for TFMPP in "party pills," while not high enough to cause adverse effects but not too low to result in inadequate detection. The pharmacokinetic parameters of BZP have already been described using this method (9).

Experimental Methods

Chemicals and Reagents

BZP (1-benzylpiperazine, 98+% purity, Lot# 11032930) was sourced from Alfa Aesar (USA). TFMPP hydrochloride (1-(α,α,α,-trifluoro-m-tolyl) piperazine, 98+% purity, Lot# 102K3678) was sourced from Sigma Aldrich (Auckland, New Zealand). 3-OH BZP (3′-hydroxy-1-benzylpiperazine) and 4-OH BZP (4′-hydroxy-1-benzylpiperazine) were custom synthesized by Sigma Aldrich. 4-OH TFMPP (1-(3-trifluoromethyl-4-hydroxyphenyl) piperazine) was kindly synthesized by Brian Palmer the Auckland Cancer Society Research Centre (University of Auckland, New Zealand). The purity of all standards was verified by LC–MS.

Methanol (>99%, spectrophotometric grade, Sigma Aldrich) and zinc sulfate monohydrate (>99%, Sigma Aldrich) were used during sample preparation. Formic acid (50%, Sigma Aldrich), ammonia solution (35%; Acros Organics; Auckland, New Zealand), LC-grade water (Millipore®, Milli-Q system; Billerica, MA), and acetonitrile (99.9%, UV grade, Sigma Aldrich) were used for the mobile phase.

Chromatographic Equipment and Conditions

Chromatography was performed using an Agilent 1100 liquid chromatography (LC) system coupled with an Agilent MSD model D single-staged quadrupole mass spectrum (MS) detector. Agilent ChemStation software (Version B 03.01) (Agilent Technologies, Goettingen, Germany) was used for data acquisition and calculations of area under the curve (AUC). Chromatographic separation was achieved using an Agilent Zorbax C18 HPLC column (4.6 mm × 150 mm, 5 μm) with a guard column (C18, 4.6 × 10, 5 μm). Chromatography was conducted at 20°C. A mobile phase of ammonium formate buffer (pH 4.5, 0.01 M, solvent A) and acetonitrile (solvent B) with a phase gradient (0–2 min 5% B; 2–5 min 10% B; 5–10 min 10–55% B; 10–12 min 55 – 5% B; 12–5 min 5% B) was used for separation. MS detection using electrospray ionization was performed. Spray chamber parameters were as follows: gas temperature 350°C (max 350°C), drying das flow rate 12 L/min (max 13.0 L/min), nebulizer pressure 35 psig (max 60 psig), voltage cap (positive and negative) 3000 V.

Detection by single-ion monitoring for each mass ion was used: m/z 177 (BZP), m/z 231 (TFMPP), m/z 193 (3-OH BZ and 4-OH BZP), m/z 247 (4-OH TFMPP). Total run time was 15 min with a flow rate of 1 mL/min and sample injection size of 20 μL. The AUC measurements of each analyte were compared to a standard curve and used for the quantification of each analyte in the samples.

Standard Solutions

Standard stock solutions of BZP, TFMPP, 3-OH BZP, 4-OH BZP, and 4-OH TFMPP were prepared in ammonium formate buffer (pH 4.5, 0.01 M) with final concentrations of 1 mg/mL. Appropriate dilutions of the stock solutions were made in formate buffer. The calibration curve samples were prepared by spiking 90 μL of human plasma with 10 μL of standard solutions, resulting in final concentrations of each analyte of 1, 5, 10, 15, 20, 25, and 50 ng/mL. The standard stock solutions and calibration curve were freshly prepared on each analysis day.

Sample Preparation

Plasma samples were deproteinized by the addition of $ZnSO_4$ (20 μL, 35%) and methanol (100 μL). Samples were vortex mixed for 1 min, centrifuged for 10 min at 11,200 × g at room temperature, and the clear supernatant was collected for analysis as described above.

Validation Procedures

The selectivity of the assay was investigated by processing and analysing blank samples prepared from six independent plasma sources. These were monitored for interfering peaks. Matrix effects were monitored by spiking independent blank plasma samples and comparing the analyte concentrations to the true spiked concentration. To determine the accuracy of the method, samples of each analyte were prepared at four concentrations (5, 10, 20, and 50 ng/mL) in each independent matrix sample and analyzed by the procedure described earlier.

Validation samples were prepared and analyzed to evaluate the intra-day and inter-day precision of the analytical method in human plasma. Precision was reported as percentage of relative standard deviation (% R.S.D.) of the estimated concentrations, and the accuracy was determined by comparing the measured concentrations of the plasma samples with the true concentration (spiked into the sample). To determine linearity, three determinations of each of the four concentrations for each analyte were performed on three consecutive days. Calibration standards were freshly prepared daily prior to analysis. Slopes, intercepts, and correlation coefficients were also determined.

Recovery of the analytes from plasma was investigated by the comparison of freshly prepared spiked buffer samples (n =3 at 5, 10, 20, and 50 ng/mL) with freshly prepared spiked plasma samples.

The short-term stability of the plasma samples was investigated by the comparison of freshly prepared spiked plasma samples (n = 3 at 5, 10, 20, and 50 ng/mL) with identical samples prepared and left at room temperature for 24 h. The freeze–thaw stability of plasma samples was investigated by the comparison of freshly prepared spiked plasma samples (n = 3 at 5, 10, 20, and 50 ng/mL) with identical samples prepared and subjected to three freeze–thaw cycles of −20°C storage for 24 h, followed by unassisted thawing at room temperature. The long-term storage stability of plasma samples was investigated by the comparison of freshly prepared spiked plasma samples (n = 3 at 5, 10, 20, and 50 ng/mL) with identical samples prepared and stored at −20°C for 3 months. Stock solution stability was investigated by the comparison of freshly prepared spiked buffer samples (50 ng/mL) with an identical sample prepared 48 h previously. Postpreparative stability was investigated by comparing two injections of a spiked plasma sample (50 ng/mL) with a 33-h delay between injections. In addition to validation protocol, trials were conducted to identify a suitable internal standard for this method.

Pharmacokinetic Trial Conditions

Study Protocol—Healthy human volunteers (18–31 years old, men n = 5; women n = 1; mean body mass index: 27.4 kg/m^2) took part in this study. To minimize risk to participants, strict exclusion criteria were used. These criteria were as follows: history of drug allergies, liver disease, diabetes, cardiovascular disease, drug or alcohol abuse, mental illness or respiratory disease. Participants on any medication were excluded from this trial.

Ethical approval to carry out this research was given by the Northern X Regional Ethics Committee of NZ (NTX06/07/080). All participants fasted for 12 h prior to drug administration and were provided with set meals 2 and 6.67 h after drug ingestion; water was also available during the course of the study. Fasting and standardized meals were used to minimize the effect of food on inter-individual variability in drug absorption from the gut. TFMPP hydrochloride (60 mg; Sigma Aldrich, New Zealand) in gelatine capsules (Size 0CS; Capsugel, West Ryde, Australia) was given as a single oral dose, and 15 blood samples were collected over a 24-h period. One blood sample was taken before drug administration ($t = 0$), and fourteen additional samples were taken postdose at $t = 15$, 30, 45, 60, 75, 90, 105, 120, 180, 240, 300, 350, 480, and 1440 min.

Total urine excreted over the 24-h test period was also collected for analysis. The total volume of urine collected for each participant was measured and recorded.

Sample Handling—Blood samples (6 mL) were collected in heparinized tubes and allowed to stand at room temperature for 30 min prior to centrifugation at 800 × g for 15 min to separate the plasma and red blood cell fractions. An aliquot of plasma (100 μL) was deproteinized by adding ZnSO4 (20 μL, 35%) and methanol (100 μL). Samples were vortex mixed for 1 min, centrifuged for 10 min at 11,200 × g, and the clear supernatant was collected for analysis as described below.

An aliquot of urine (500 μL) was incubated at 37°C for 12 h with a mixture (100 μL, 100,000 Fishman units per mL) of β-glucuronidase (EC 3.2.1.31, Sigma, USA) and aryl sulfatase (3.1.6.1, Sigma, USA) to hydrolyze any conjugates. The urine sample was centrifuged for 10 min at 11,200 × g, and the supernatant was analyzed as described below. A second aliquot (200 μL) of urine was filtered (0.45-μm RC-membrane syringe filter; Sartorius, NSW, Australia) and analyzed directly as described below to determine the relative excretion of conjugated metabolites.

Instrumental Analysis—Plasma and urine samples were analyzed using the method described above. Peaks yielded by single ion monitoring (SIM) were used for quantification, while retention time and mass spectra of peaks in the total ion chromatogram (TIC) were compared to the standards to confirm the identity of the peaks in the SIM traces. The presence of additional metabolites [as proposed by Staack et al. (3)] was investigated by searching for the mass ions and major ions, but these were not found in the plasma samples. Quantification: Standard curves of plasma samples spiked with TFMPP and 4-OH TFMPP (5, 10, 15, 20, 25, and 50 ng/mL; $n = 3$) were prepared and used to determine the concentration of these compounds in the plasma samples. The area under the curve for TFMPP and 4-OH TFMPP was measured (using Agilent ChemStation software) and compared to standard curves. Similar standard curves were prepared for TFMPP and 4-OH TFMPP in urine and cleaned up by the enzyme hydrolysis and the microfiltration methods.

Data Analysis—WinNonLin software (by Pharsight) was used for modeling and analysing the pharmacokinetic parameters. A two-compartment model, with first-order absorption, a lag time, and first-order elimination, was used for this data. Primary pharmacokinetic parameters were calculated, including apparent clearance (Cl/F), volume of distribution for the central compartment (V1/F), inter-compartment clearance (Q/F), and volume of distribution for the peripheral compartment (V2/F), and secondary parameters, including rate constants, were derived from these values.

Results

Method Validation

The specificity and selectivity of the method is demonstrated by the comparison of the chromatogram of a blank human plasma sample with that of plasma spiked with each compound (20 ng/mL) (Fig. 1). The retention times for BZP, TFMPP, 3-OH BZP, 4-OH BZP, and 4-OH TFMPP were 7.61, 9.16, 2.46, 3.55, and 7.80, respectively.

No significant interfering peaks were detected in any of the six independent blank plasma samples. Accuracy of quantification for this method was found to be >90% for all five compounds between 5 and 50 ng/mL (Table 1). The results from Table 1 also demonstrate that no significant matrix effects are evident in the six independent plasma samples used.

Less than 10% variation was observed between samples run on the same day (intra-day) and on consecutive days (inter-day) (Tables 2 and 3, respectively). These measures of accuracy and precision were less than the 15% limit of variation set by the FDA method validation guidelines and hence suitable for a bioanalytical method.

Linearity of the concentration–response relationship was verified for this method. This was achieved using the slopes and coefficients of determination (R^2) of nine linear standard curves generated on three consecutive days (Table 4).

Recovery of analytes was found to be greater that 85% for all analytes (Table 5). Short-term stability of the analytes was found to be >80% for 4-OH BZP, >90% for 3-OH BZP and TFMPP, and >95% for BZP and 4-OH TFMPP (Table 6). Freeze–thaw stability of the analytes was found to be >80% for 3-OH BZP and 4-OH BZP, >90% for BZP, and >95% for TFMPP and 4-OH TFMPP (Table 7). Long-term stability of the analytes was found to be >80% for 3-OH BZP, >85% for 4-OH BZP, and >90% for BZP, TFMPP, and 4-OH TFMPP (Table 8). Stability of the stock solution was found to be >95% for all analytes (Table 9). Postpreparative stability of a plasma sample was found to be greater that 95% for all analytes (Table 10). The lower limit of quantification was 5 ng/mL for all analytes.

TFMPP pharmacokinetics

TFMPP—The average concentration of TFMPP ($n = 6$) in plasma was plotted over time (Fig. 1). The results demonstrate a mean time delay (Tlag) of 30 min before TFMPP was detected. TFMPP was undetectable in all participants at 15 min and all but two participants at 30 min but at concentrations below limit of quantitation (LOQ). The peak plasma concentration (C_{max}) of 24.1 ng/mL (±2.1 ng/mL; SEM) was reached 90 min (1.50 h) (T_{max}) postdose. The absorption half-life was calculated to be 24.6 min (±7.9 min). At the 480-min (8-h) sample, all participants had quantifiable amounts of TFMPP. By the end of the 24-h sampling period, TFMPP concentrations had dropped below LOQ. The apparent clearance (reported here as clearance/bioavailability or Cl/F) of TFMPP was 384.24 L/h (±45.06 L/h). Table 11 presents the primary pharmacokinetic parameters from this data.

As demonstrated in the logarithmic plot in Fig. 2, the clearance phase was divisible into two parts—an early phase of rapid clearance followed by a slower phase. Half-lives ($t_{1/2}$) were calculated for each phase of the elimination curve. Between 90 min (T_{max}) and 2 h, a rapid clearance with an average $t_{1/2}$ of 2.04 h (±0.19 h) can be observed, followed by a period of much slower clearance with an average $t_{1/2}$ of 5.95 h (±1.63 h) between 3 and 8 h after the dose.

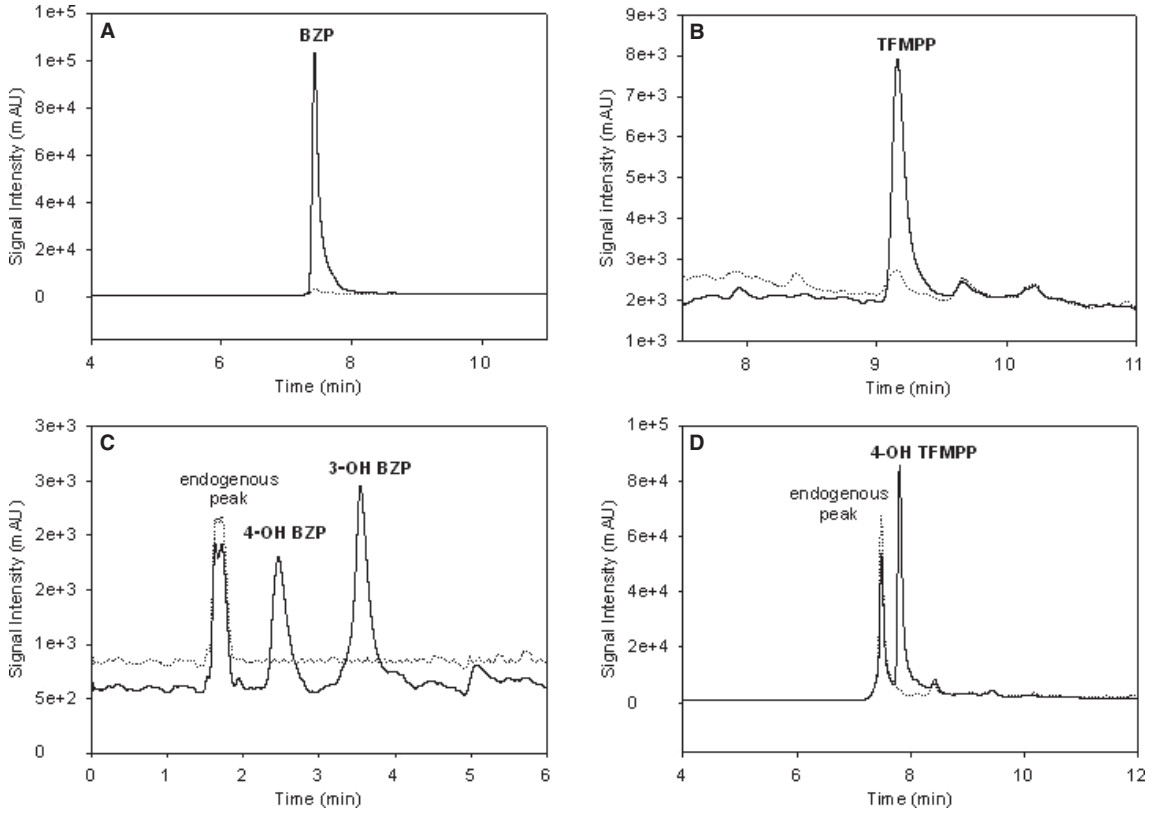


FIG. 1—*Blank and spiked plasma with (clockwise from top left) benzylpiperazine (BZP), TFMP, 4-OH TFMPP and 3-OH BZP and 4-OH BZP.*

TABLE 1—*Variation between spiked plasma samples* (n = *6): a measure of accuracy and matrix effects.*

	3-OH Benzylpiperazine (BZP)		4-OH BZP		BZP		4-OH TFMPP		TFMPP	
Spike (ng/mL)	%Acc	Range	%Acc	Range	%Acc	Range	%Acc	Range	%Acc	Range
5	93.60	4.47–5.05	92.22	4.60–4.75	90.37	4.38–4.90	91.67	4.49–4.95	94.17	4.62–5.05
10	91.98	8.72–10.11	90.16	11.20–13.74	91.20	9.81–11.44	96.60	8.96–10.78	91.62	9.88–11.20
15	96.07	14.40–15.30	91.69	14.50–17.36	91.30	15.04–16.99	90.18	13.16–14.76	95.07	13.73–16.19
20	91.16	20.92–22.04	96.68	22.01–23.43	94.36	18.41–19.53	90.93	18.84–25.87	98.23	18.89–20.49
50	98.87	47.90–52.45	97.08	47.51–50.07	98.07	48.18–54.43	92.53	41.29–51.29	95.52	51.06–52.65

TABLE 2—*Intra-day variation between spiked plasma samples* (n = *3).*

	3-OH Benzylpiperazine (BZP)			4-OH BZP			BZP			4-OH TFMPP			TFMPP		
Spike (ng/mL)	Mean	SD	RSD	Mean	SD	RSD	Mean	SD	RSD	Mean	SD	RSD	Mean	SD	R.S.D
5	4.68	0.07	1.41	6.02	0.58	9.59	5097	0.07	11.11	5.01	0.06	1.21	5.07	0.34	6.75
10	9.30	0.24	2.59	12.05	1.16	9.59	11.94	0.13	11.11	10.02	0.12	1.21	10.15	0.69	6.75
15	14.48	0.63	4.37	15.31	1.32	862	15.59	0.39	2.52	13.88	0.05	0.35	15.05	0.59	3.91
20	21.18	0.39	1.85	21.82	1.18	5.40	18.47	0.68	3.68	21.00	1.22	5.80	20.31	0.57	2.80
50	49.51	0.48	0.97	48.31	0.62	1.29	51.26	3.83	7.48	48.39	2.61	5.40	51.52	1.11	2.16

4-OH TFMPP—A single metabolite, 4-OH TFMPP, was detected in plasma with a mean C_{max} of 20.2 ng/mL (±4.6 ng/mL) with a T_{max} of 90 min and Tlag of 30 min (Fig. 3). This peak was identified as 4-OH TFMPP by comparison of retention time and spectroscopy data to an authentic standard. 4-OH TFMPP was undetectable in all participants at 15 min and all but two participants at 30 min but at concentrations below LOQ. 4-OH TFMPP was detectable in all participants at 350 and 480 min but were below LOQ in all but one participant at 350 and 480 min.

The elimination period demonstrates a constant decline from 105 min (T_{max}) to 5 h (after which more data points fall below LOQ). The elimination rate was found to be linear (R^2 = 0.971), and an average $t_{1/2}$ of 6.57 h (±0.60 h) was observed for this metabolite.

TABLE 3—*Variation between three sets of plasma samples analyzed on three consecutive days.*

Spike (ng/mL)	Day	3-OH Benzylpiperazine (BZP)			4-OH BZP			BZP			4-OH TFMPP			TFMPP		
		Mean	SD	RSD	Mean	SD	RSD	Mean	SD	RSD	Mean	SD	RSD	Mean	SD	RSD
5	1	4.61	0.11	2.39	5.36	0.25	4.68	5.94	0.04	0.72	4.97	0.04	0.90	4.68	0.41	8.85
	2	4.70	0.15	3.19	6.35	0.55	8.63	6.04	0.36	5.92	5.08	0.06	1.18	5.26	0.11	2.13
	3	4.74	0.32	6.76	6.36	0.40	6.29	5.92	0.53	8.96	4.97	0.21	4.15	5.28	0.36	6.76
10	1	9.03	0.21	2.37	10.71	0.24	2.24	11.87	0.35	2.96	9.94	0.40	4.02	9.36	0.41	4.43
	2	9.41	0.14	1.44	12.71	1.05	8.30	12.09	1.17	9.71	10.15	0.55	5.40	10.52	0.11	1.07
	3	9.47	0.43	4.58	12.72	1.24	9.74	11.85	1.04	8.81	9.95	0.64	6.47	10.56	0.36	3.38
15	1	13.75	1.04	7.58	13.43	0.78	5.81	15.73	0.04	0.27	13.87	0.39	2.78	15.73	1.07	6.78
	2	14.85	0.25	1.67	16.20	1.55	9.56	15.14	1.36	8.96	13.84	0.20	1.48	14.74	0.62	4.19
	3	14.84	0.01	0.07	17.36	0.40	2.30	15.89	0.53	3.34	13.93	0.39	2.81	14.69	0.89	6.08
20	1	20.80	0.05	0.23	20.47	0.90	4.41	17.72	0.04	0.25	19.59	1.58	8.08	20.95	0.13	0.61
	2	21.16	0.03	0.14	22.39	1.40	6.23	18.61	0.04	0.24	20.67	1.45	7.03	20.15	1.10	5.48
	3	21.59	0.39	1.83	22.61	0.51	2.26	19.06	0.41	2.13	20.32	1.32	6.52	19.84	0.66	3.35
50	1	49.22	1.01	2.05	47.90	0.41	0.85	55.43	0.51	0.92	51.40	1.92	3.73	50.31	1.44	2.87
	2	49.24	1.04	2.11	48.01	0.04	0.07	44.69	2.50	5.59	47.04	4.46	9.48	52.49	0.63	1.20
	3	50.07	1.78	3.56	49.03	0.90	1.84	50.47	2.96	5.87	46.73	4.62	9.88	51.76	0.32	0.62

TABLE 4—*Linearity: Comparison of slope and* R^2 *of nine linear calibration curves.*

	3-OH Benzylpiperazine (BZP)		4-OH BZP		BZP		4-OH TFMPP		TFMPP	
	Slope	R^2	Slope	R^2	Slope	R^2	Slope	R^2	Slope	R^2
Day 1	0.9814	0.9960	0.9730	0.9961	1.0695	0.9709	0.9700	0.9936	0.9644	0.9943
	1.0184	0.9988	0.9727	0.9948	1.0641	0.9552	1.0448	0.9854	0.9856	0.9853
	0.9802	0.9986	0.9548	0.9933	0.8596	0.9799	1.0041	0.9806	1.0166	0.9950
Day 2	0.9943	0.9916	0.9647	0.9935	0.9160	0.9763	1.0265	0.9911	1.0017	0.9762
	1.0293	0.9959	0.9440	0.9902	0.8823	0.9894	0.8549	0.9937	1.0441	0.9922
	0.9993	0.9928	0.9534	0.9797	0.9476	0.9651	0.9212	0.9986	1.0023	0.9728
Day 3	1.0210	0.9953	0.9913	0.9952	1.0358	0.9520	1.0110	0.9972	0.9890	0.9743
	1.0611	0.9960	0.9750	0.9901	0.9348	0.9903	0.8281	0.9910	1.0200	0.9969
	0.9941	0.9871	0.9460	0.9851	0.9238	0.9723	0.9475	0.9956	1.0041	0.9771
Mean	1.0088	0.9947	0.9639	0.9909	0.9593	0.9724	0.9565	0.9919	1.0031	0.9849
SD	0.0262	0.0037	0.0156	0.0054	0.0908	0.0135	0.0760	0.0057	0.0228	0.0099
RSD	2.60	0.37	1.62	0.55	9.46	1.38	7.95	0.58	2.27	1.00

TABLE 5—*Percentage of analyte recovered from spiked plasma samples (compared to spiked buffer samples).*

Spike (ng/mL)	3-OH Benzylpiperazine (BZP)	4-OH BZP	BZP	4-OH TFMPP	TFMPP
5	88.64	93.93	94.61	94.01	90.36
10	88.95	100.59	92.50	93.71	79.54
20	89.65	93.87	99.82	93.77	96.56
50	87.33	87.33	91.51	94.54	94.99

TABLE 6—*Percentage of analyte recovered from spiked plasma samples after storage at room temperature for 24 h.*

Spike (ng/mL)	3-OH Benzylpiperazine (BZP)	4-OH BZP	BZP	4-OH TFMPP	TFMPP
5	94.68	89.40	98.53	97.33	96.53
10	97.52	83.80	99.71	97.32	99.33
20	94.61	91.73	98.43	96.27	97.17
50	91.91	92.68	97.46	98.40	93.10

Urine Samples—TFMPP and 4-OH TFMPP were measured in neat 24-h urine samples in a 6:1 ratio or 73.7 μg (±19.3 μg) of TFMPP versus 11.7 μg (±3.59 μg) of 4-OH TFMPP excreted in 24 h. Enzymatic hydrolysis of the urine samples resulted in higher amounts of these analytes (255.8 ± 34.7 μg of TFMPP vs. 110.4 ± 15.6 μg of 4-OH TFMPP). However, efficiency of the hydrolysis method is not known (because of the absence of standards for the conjugates), thus the total amount of drug excreted cannot be accurately quantified. The total amount of drug quantified in the hydrolyzed urine was 359 μg (±50 μg) or 0.6% of the dose.

The discrepancy between unconjugated and total TFMPP and 4-OH TFMPP concentrations suggests the presence of the glucuronide and sulfate conjugates (70% of TFMPP and 90% of 4-OH TFMPP). The presence of these conjugates was investigated by scanning the TIC of the neat urine samples for their mass ions (422 m/z for *O*- or *N*-glucuronide of 4-OH TFMPP, 327 m/z for *O*- or *N*-sulfate of 4-OH TFMPP, 406 m/z for *N*-glucuronide of TFMPP and 310 m/z for *N*-sulfate of TFMPP) and for other distinguishing ion fragments (including 231 and 247 m/z). A peak at 11.7 min with a mass ion and spectrum consistent with *N*-glucuronide TFMPP was detected following analysis of the mass spectra (Fig. 4). The identification of this metabolite is speculative as standards of the conjugates were not available, hydrolysis of glucuronides and sulfates was conducted simultaneously (therefore not allowing for the distinction between the two), and no further

TABLE 7—*Percentage of analyte recovered from spiked plasma samples after three freeze–thaw cycles.*

Spike (ng/mL)	3-OH Benzylpiperazine (BZP)	4-OH BZP	BZP	4-OH TFMPP	TFMPP
5	85.15	84.43	95.35	96.64	97.16
10	87.50	83.47	93.71	95.64	97.24
20	82.17	84.30	94.97	96.52	96.34
50	85.79	85.52	97.37	97.76	97.90

TABLE 8—*Percentage of analyte recovered from spiked plasma samples after long-term storage.*

Spike (ng/mL)	3-OH BZP	4-OH BZP	BZP	4-OH TFMPP	TFMPP
5	90.28	90.95	96.42	94.17	95.74
10	84.31	88.44	93.35	87.83	96.94
20	94.61	91.73	98.43	96.27	97.17
50	91.91	92.68	97.46	98.40	93.10

BZP, Benzylpiperazine.

TABLE 9—*Recovery of analyte from a stock solution stored at room temperature for 24 h.*

3-OH Benzylpiperazine (BZP)	4-OH BZP	BZP	4-OH TFMPP	TFMPP
97.17	99.49	98.64	98.73	98.87

TABLE 10—*Recovery of analyte from a plasma sample in an LC–MS autosampler (5°C) for 33 h.*

3-OH Benzylpiperazine (BZP)	4-OH BZP	BZP	4-OH TFMPP	TFMPP
98.81	97.10	96.31	99.97	98.95

TABLE 11—*Pharmacokinetic parameters (±SEM) for TFMPP (60 mg).*

Apparent clearance; Cl/F (L/h)	384.24 (±45.06)
Apparent volume of distribution; V/F (L)	891.34 (±176.39)
Inter-compartmental clearance; Q/F (L/h)	1015.22 (±180.26)
Volume of peripheral compartment; V2/F (L)	2201.97 (±409.71)

analyses were conducted to confirm the identity of this peak. No conclusive evidence was found for the other conjugates, although hydrolysis data indicate the presence of a conjugate of 4-OH TFMPP (Fig. 5).

Research by Staack and colleagues proposed that TFMPP had a number of possible metabolites (3); however, only 4-OH TFMPP could be detected and quantified in these plasma and urine samples. The conjugated metabolite *N*-glucuronide TFMPP was only tentatively identified in the urine samples. Also, further searching for the mass ions of other proposed metabolites within the LC–MS trace of both plasma and urine samples was unsuccessful. This may be the result of poor sensitivity of the method for these analytes. Another explanation may be that not all of the metabolites discussed by Staack were formed at the dose given (about 50% of the maximum dose used recreationally).

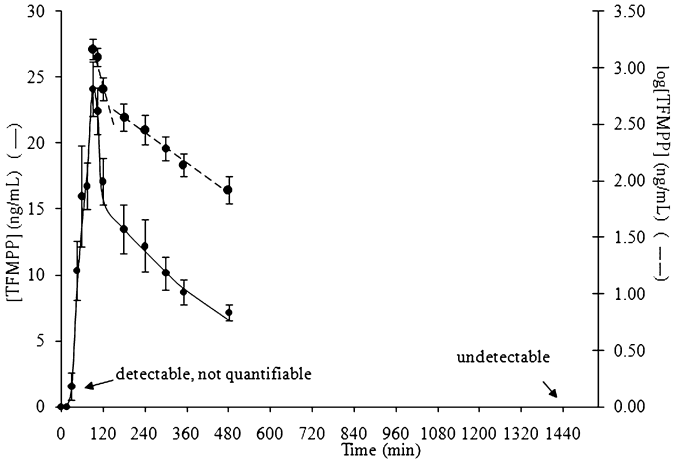


FIG. 2—*Mean plasma concentrations of TFMPP (±SEM) showing two phases of elimination over 24 h following a single 60 mg oral dose* (n = 6).

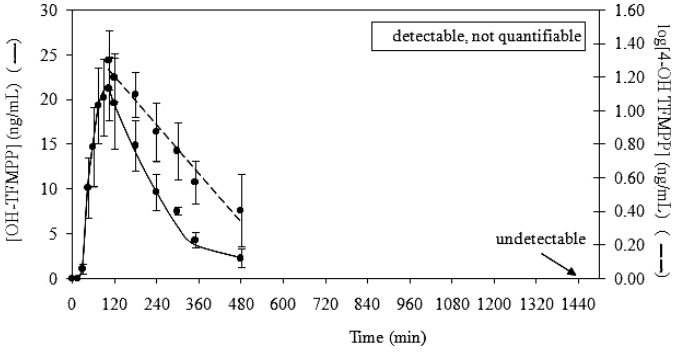


FIG. 3—*Mean plasma concentrations of 4-OH TFMPP (±SEM) over 24 h following a single 60 mg oral dose of TFMPP* (n = 6).

Discussion

Method Validation

The reproducibility of this method is found to comply with limits set by industry guidelines (8) including specificity, selectivity, accuracy, precision, linearity, and stability. As the sample preparation involved minimal sample handling and no phase transfer steps, a major source of error was also eliminated. Advances in chromatographic systems have greatly reduced the amounts of error associated with analysis techniques.

Matrix effects can adversely impact on the validity of bioanalytical methods. Endogenous compounds in plasma, such as phospholipids, while not necessarily being visible in the chromatogram themselves, are a significant source of ion suppression of analytes introducing error into the analytical method and increasing the chances of false negatives at low analyte concentrations (10). Matrix effects were not found to be a cause for concern in this method as determined by the high accuracy of this method across a number of independent plasma samples. However, as this method has only been validated for plasma samples, interference from other matrices cannot be discounted at this stage. Using whole blood or urine samples may indeed produce significant matrix effects because of the presence of different endogenous compounds. Furthermore, biological samples can demonstrate substantial variability in the type and amount of interferents in the same individual at different times.

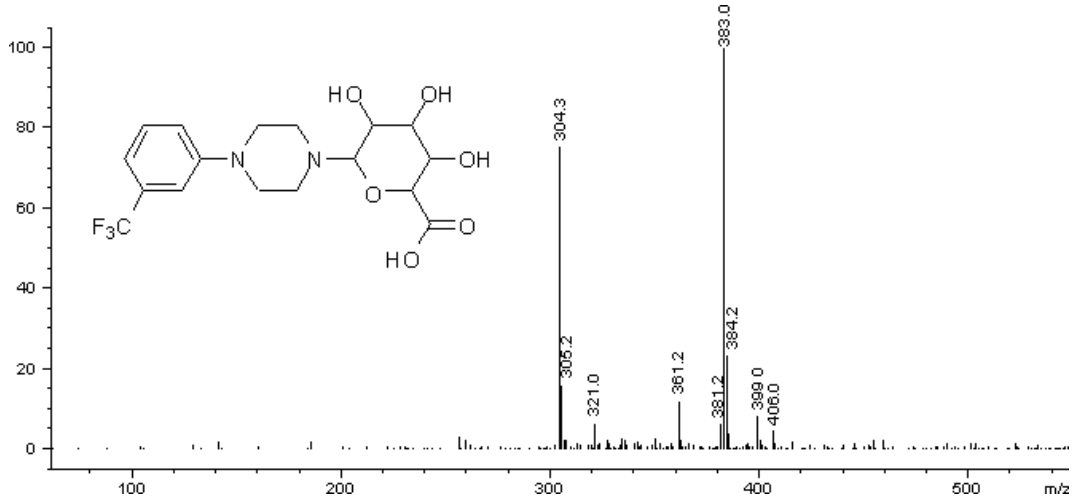


FIG. 4—*Mass spectrum of urinary metabolite* N*-glucuronide TFMPP.*

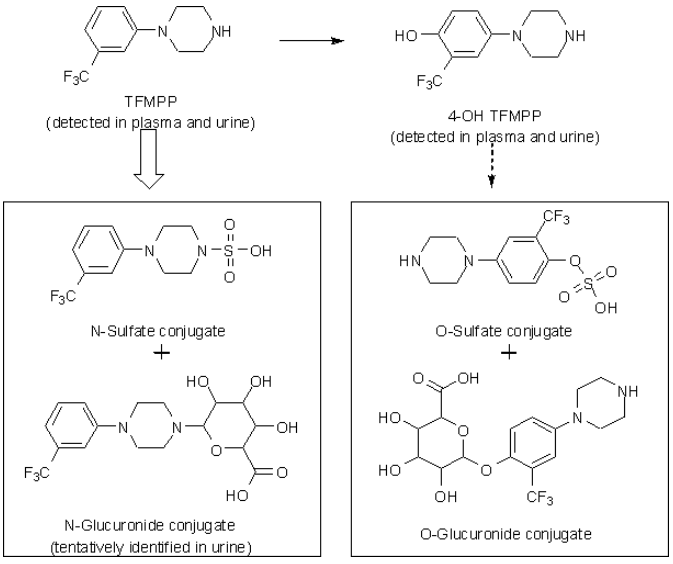


FIG. 5—*A schematic representation of the metabolites of TFMPP detected in plasma and urine, and other potential conjugated metabolites.*

An issue faced in the development of this method was identifying a suitable internal standard. A number of related piperazines were trialed as internal standards, including MeOPP, pFPP, and mCPP, but their interference with the chromatography and detection of the analytes precluded their use in this method. These compounds not only co-eluted with the drugs of interest but also had similar masses and fragmentation patterns to the metabolites of BZP. Using deuterated standards of the analytes may provide a suitable internal standard for this method; however, these were not available at the time of method development and validation. As this method demonstrates a low error rate and allows for the simultaneous detection of five analytes within a short run time, for the purposes of the pharmacokinetic analyses of the plasma samples, this method is adequate despite the omission of an internal standard. Nevertheless, as internal standards also correct any alteration in signal because of matrix effects (11), for forensic detection of these drugs, especially in a range of matrix types, the inclusion of deuterated standards will provide a more rigorous analytical method.

TFMPP Pharmacokinetics

An obvious feature in the plasma profiles of TFMPP is the presence of a lag time of at least 30 min prior to either compound being detected. All participants demonstrated undetectable levels of TFMPP in the 0-, 15-, and 30-min samples. This suggests that, following an oral dose, TFMPP absorption from the gut into the bloodstream is not immediate. This lag can have a bearing on the onset of effects and, as TFMPP is self-administered recreationally, may lead to repeated dosing. However, this lag phase may be an artifact of the fasting period leading to slower stomach emptying and may not be a true representation of TFMPP absorption under "normal" circumstances where food (and alcohol) might be consumed prior to taking the dose.

The concentrations of TFMPP in plasma clearly show that there are two distinct disposition phases following C_{max}, a finding consistent for all participants. The early phase demonstrates a sharp decrease ($t_{1/2} \sim 2$ h) between 90 min and 2 h after drug administration. The rapid decrease in the early disposition phase is likely because of redistribution of the drug into organs such as the kidney and liver, which are responsible for elimination or alternatively sites of action such as the brain. After this distribution period, the plasma concentration then decreases at a slower elimination rate which is clearly demonstrated in the logarithmic plot in Fig. 2 and is clearly distinguishable from the earlier and more rapid phase. It is important to note that the lack of samples between 480 and 1440 min (8 and 24 h) is a limitation of this data, which may obscure a further elimination phase and affect the robustness of the half-life calculation for the terminal phase as a result.

The initial distribution phase also raises the issue of the tissue concentrations of TFMPP in other organ systems. The results of this study demonstrate that peak plasma concentrations of TFMPP were found 90 min after dosing and that little or no TFMPP remains in the plasma after 24 h. However, it is simplistic to assume that the plasma T_{max} is representative of the time at which the peak pharmacodynamic effects of TFMPP manifest. For example, in the case of methamphetamine, plasma concentrations do not reflect those found in the brain (12). TFMPP may exhibit a similar disparity between plasma and brain tissue concentrations. The fact that this drug has observable effects in the brain (5) suggests that TFMPP crosses the blood–brain barrier.

4-OH TFMPP concentrations were lower than those of TFMPP at all times, consistent with previous studies demonstrating that TFMPP is only partially metabolized (1–3). The concentration of 4-OH TFMPP falls below the limit of detection by 24 h. Urine samples also demonstrate low concentrations of 4-OH TFMPP compared to TFMPP.

The analytical method is validated for detecting and quantifying TFMPP and 4-OH TFMPP in blood plasma but not in urine. Furthermore, no standards were present for the conjugated metabolites of TFMPP. Therefore, only tentative results may be gained from the urine samples.

This research has shown that TFMPP in urine is excreted primarily in its conjugated form, probably as an *N*-glucuronide that was tentatively identified in the filtrate of urine samples, accounting for around 70% of the excreted TFMPP. The hydroxylated metabolite 4-OH TFMPP is also excreted mainly as a conjugate (around 90%), although the identity of this conjugate has not been determined.

The total amount TFMPP (including metabolites) excreted in the urine accounts for <1% of the dose administered, suggesting that TFMPP has low bioavailability, unless poor hydrolysis efficiency, other routes of excretion (such as biliary excretion) or strong tissue or protein binding account for the low recovery of TFMPP.

The suitability of urine for detection TFMPP exposure cannot be adequately concluded from the study design. Had individual urine specimens been collected, it may have been found that, for a short window of time after administration, TFMPP or 4-OH TFMPP could be readily detected in urine. Observed concentrations in 24-h pooled urine are likely to be diluted by negative specimens collected immediately after oral administration (when the drug has not yet had time to be absorbed, metabolized or be excreted) or specimens toward the end of the collection period (when all drug residuals may have been eliminated).

We report for the first time, plasma kinetics of TFMPP in humans. Further studies are required to demonstrate tissue uptake into the brain. This research may influence the design of future pharmacological studies investigating TFMPP.

Conclusion

An HPLC method with MS detection has been developed for quantification of BZP, TFMPP, and their metabolites 3-OH BZP, 4-OH BZP, and 4-OH TFMPP in human plasma. All results were within the acceptable ranges for bio-analytical purposes. The chromatography and detection methods allowed clear separation of these compounds, with no interfering peaks, and no detectable matrix effects. The assay was validated, and the results demonstrate that the standard curve for each analyte is linear over the concentration range investigated. While not used in this method, there is merit in finding a suitable internal standard for forensic detection. This assay is reproducible and accurate, can resolve mixtures of these analytes, has an acceptable analysis time, and can be applied successfully to quantification of these analytes in plasma samples down to concentrations of 5 ng/mL. This method has been successfully applied to the analysis of BZP, TFMPP, and their hydroxylated metabolites in human plasma.

Acknowledgments

Sincere thanks to Dr. Brian Palmer (Auckland Cancer Society Research Centre, University of Auckland) for the custom synthesis of 4-OH TFMPP for this paper.

References

1. Antia U, Tingle M, Russell B. Metabolic interactions with piperazine-based 'party pill' drugs. J Pharm Pharmacol 2009;61(7):877–82.
2. Staack R, Fritschi G, Maurer H. Studies on the metabolism and toxicological detection of the new designer drug N-benzylpiperazine in urine using gas chromatography- mass spectrometry. J Chromatogr B Analyt Technol Biomed Life Sci 2002;773:35–46.
3. Staack R, Paul LD, Springer D, Kraemer T, Maurer HH. Cytochrome P450 dependent metabolism of the new designer drug 1-(3-trifluoromethylphenyl) piperazine (TFMPP): in vivo studies in Wistar and Dark Agouti rats as well as in vitro studies in human liver microsomes. Biochem Pharmacol 2004;67:235–44.
4. Gordon M, Millar J, Kirk I, Lim V, Waldie K, Russell B, editors. The effects of Benzylpiperazine (BZP) administration on the Poffenberger Paradigm of Interhemispheric Transfer in healthy adult right handed males. Proceedings of the International Australasian Winter Conference on Brain Research; 2007 Aug 25–29; Queenstown, New Zealand, http://psy.otago.ac.nz/awcbr/Abstracts/Abstracts2007.
5. Lee H, Gordon M, Millar J, Kirk I, Lim V, Russell B, editors. The acute effects of Trifluoromethylphenylpiperazine (TFMPP) administration on the amplitude of the P300 in healthy adult right handed human males. Proceedings of the International Australasian Winter Conference on Brain Research; 2007 Aug 25-29; Queenstown, New Zealand, http://psy.otago.ac.nz/awcbr/Abstracts/Abstracts2007.htm.
6. Russell B, Gordon M, Millar J, Kirk I, Lim V, Waldie K, editors. Benzylpiperazine (BZP) decreases the amplitude of the P300 Event-Related Potential (ERP) in healthy adult right-handed human males. Proceedings of the International Australasian Winter Conference on Brain Research; 2007 Aug 25–29; Queenstown, New Zealand, http://psy.otago.ac.nz/awcbr/Abstracts/Abstracts2007.html.
7. Murphy M, Antia U, Chang H, Han J, Ibrahim A, Russell B, et al. Party Pills and Drug–Drug interactions. N Z Med J 2009;122(1293):16–25.
8. U.S Department of Health and Human Services, Food and Drug Administration, Centre for Drug Evaluation and Research, Centre for Veterinary Medicine. Guidance for industry: bioanalytical method validation, May 2001, http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM070107.pdf.
9. Antia U, Lee H, Kydd R, Tingle M, Russell B. Pharmacokinetics of 'party pill' drug N-benzylpiperazine (BZP) in healthy human participants. Forensic Sci Int 2009;186(1):63–7.
10. Chambers E, Wagrowski-Diehl DM, Lu Z, Mazzeo JR. Systematic and comprehensive strategy for reducing matrix effects in LC/MS/MS analyses. J Chromatogr B Analyt Technol Biomed Life Sci 2007;852(1–2):22–34.
11. Van EeckhautA, Lanckmans K, Sarre S, Smolders I, Michotte Y. Validation of bioanalytical LC-MS/MS assays: evaluation of matrix effects. J Chromatogr B Analyt Technol Biomed Life Sci 2009;877(23):2198–207.
12. Riviere G, Gentry W, Owens S. Disposition of methamphetamine and its metabolite amphetamine in brain and other tissues in rats after intravenous administration. J Pharmacol Exp Ther 2000;292(3):1042–7.

Additional information and reprint requests:
Ushtana Antia, M.Sc.
School of Pharmacy
University of Auckland
Private Bag 92019
Auckland
New Zealand
E-mail: u.antia@auckland.ac.nz

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01408.x
Available online at: interscience.wiley.com

PAPER

TOXICOLOGY; PATHOLOGY AND BIOLOGY

William J. Minett,[1,2] *M.S.; Tara L. Moore,*[2] *Ph.D.; Matthew P. Juhascik,*[3] *Ph.D.; Henry M. Nields,*[1] *M.D., Ph.D.; and Mindy J. Hull,*[1] *M.D.*

Concentrations of Opiates and Psychotropic Agents in Polydrug Overdoses: A Surprising Correlation Between Morphine and Antidepressants

ABSTRACT: The relationship between postmortem concentrations of morphine and co-detected psychoactive drugs in fatal overdoses is examined. Morphine and other drugs were detected in 161 medicolegal autopsy cases. Subsets of these morphine-positive cases based on drug class were established, including opioids, antidepressants, ethanol, benzodiazepines, and "other." Each subset was split into high or low concentration groups based on median drug concentrations. Morphine concentrations of the [high] and [low] groups were compared, with no significant difference in morphine concentration identified in the opioid, ethanol, or benzodiazepine subsets. The "other" drug class was too heterogeneous for statistical assessment. Morphine concentrations did show a significant direct relationship ($p = 0.01$) with antidepressants, namely increased concentrations of antidepressant drugs are associated with an increased concentration of morphine. This trend probably remains even after excluding cocaine-positive cases. The unsuspected finding that postmortem concentrations of antidepressants positively correlate with morphine levels may be important in the treatment of depression in drug addicts.

KEYWORDS: forensic science, forensic pathology, toxicology, opiates, antidepressants, drug overdose

A study by the Boston Public Health Commission has found that deaths from drugs and alcohol have risen dramatically from 2005 to 2006 mainly because of an increase in inexpensive heroin and the growing addiction to prescription medications. In 2006, 176 people died from substance abuse in Boston, MA, reflecting a 32% increase in deaths since 2005. Substance abuse is now considered the fifth leading cause of death (with a rate of 176 deaths per 100,000 population) among Bostonians behind cancer, heart disease, injuries, and stroke, respectively (1). By comparison, a 2003 study by the Drug Awareness Warning Network of six participating states (Maine, New Hampshire, Vermont, Maryland, Utah, and New Mexico) found opiate misuse fatality rates between 7.2 and 11.6 per 100,000 people (2).

A suspected cause for the increasing fatal overdoses is a resurgence of the popularity of heroin throughout New England since the late 1990s, presumably because of its low price (3). Furthermore, opiates were not commonly the sole factor in death; rather the deaths more often resulted from a combination with other drugs (2). Morphine is rarely the only drug detected in autopsy (4,5). This polydrug use may be a result of an individual's substance abuse "career," with drugs used in early stages carried through and used along with the drugs of later stages (6). The major drugs associated with heroin overdose include alcohol, benzodiazepines, and antidepressants (7). Combinations with additional respiratory depressants are of particular risk (8).

Although the phenomenon of fatal polydrug overdoses is widely recognized, few studies have investigated the importance of the postmortem concentrations of individually involved drugs. It is widely held that concurrent opiate and other central nervous system depressant drug use is more dangerous than opiate use alone. In this study, medicolegal autopsy cases of polydrug overdoses that include an opiate as well as another psychotropic drug are examined. This article proposes that the postmortem concentration of morphine will be indirectly proportional to the concentration of the coadministered central nervous system depressant drug. In other words, in polydrug overdoses in which high concentrations of a central nervous system depressant drug are detected, we expect morphine levels to be comparatively lower, and vice versa.

Methods

All postmortem examinations with toxicologic analysis performed by the Commonwealth of Massachusetts Office of the Chief Medical Examiner (MA OCME) (Boston, MA) between January 1, 2006 and December 31, 2006 in which the term "opiate" was a part of the cause of death were investigated for possible inclusion in this study. By convention in our practice, overdoses involving morphine (or heroin) are usually certified "acute opiate intoxication" or "acute intoxication because of the combined effects of opiates [and other drugs]." A chart review was performed for each identified case, with abstraction of the following

[1]Office of the Chief Medical Examiner, Commonwealth of Massachusetts, Boston, MA.
[2]Department of Anatomy and Neurobiology, Boston University School of Medicine, Boston, MA.
[3]UMass Memorial Forensic Toxicology Laboratory, Worcester, MA.

Received 26 Jan. 2009; and in revised form 29 June 2009; accepted 3 July 2009.

information for each decedent: age, height, weight, sex, manner of death, cause of death, whether drug paraphernalia was present at the scene, and other unique circumstances. Toxicologic data, including biological matrix and qualitative/quantitative drug analyses, was also determined. Ultimately, to be included in this study, an opiate (morphine, heroin, 6-monoacetylmorphine, or codeine) along with at least one other psychotropic drug had to be detected in a blood matrix. A determination was then made as to whether the opiate use was licit or illicit, with only cases of illicit use included for analysis.

Toxicologic analysis for all medicolegal postmortem examinations conducted by the MA OCME was performed by the University of Massachusetts Memorial Medical Center Forensic Toxicology Laboratory. As previously described (9), multiple specimen types per postmortem examination were submitted as appropriate and may have included: heart blood, femoral blood, urine, vitreous humor, and other tissues and fluids. The laboratory's analytical scheme includes: presumptive six drug/drug class blood analysis by enzyme-linked immunosorbent assay (Venture Labs, Redwood City, CA) for benzodiazepines, cocaine, fentanyl, methadone, opiates, and oxycodone; presumptive 10 drug/drug class urine screening by enzyme multiplied immunoassay technique (EMITÒ II Plus; Dade-Behring, Deerfield, IL) for amphetamines, barbiturates, benzodiazepines, cannabinoids, cocaine, 3,4-methylenedioxymethamphetamine (MDMA), methadone, opiates, phencyclidine, and propoxyphene; an alkaline extractable drug screen in blood by gas chromatography/mass spectrometry (GC/MS); and volatile analyses for methanol, ethanol, acetone, and isopropanol by headspace dual column GC/flame ionization detection. Confirmation and quantitation for (free) morphine was performed by GC/MS-selected-ion monitoring using a deuterated internal standard following solid-phase extraction and derivatization with *N,O*-bis(trimethylsilyl)trifluoroacetamide with 1% trimethylchlorosilane (BSTFA + 1% TMCS). The lower and upper limits of quantitation for morphine are 20 ng/mL and 2000 ng/mL, respectively. All other drug analyses were performed either in house when possible or at a national reference laboratory. The preferred analytical algorithm is presumptive screening in heart blood and confirmatory quantitative analyses in a single unpooled blood source, generally peripheral blood when available.

Cases were subsequently divided into subsets based on, along with opiates, the codetection of other psychotropic drugs classes, namely benzodiazepines, ethanol, opioids, antidepressants, and "other" (Table 1). Each drug class subset was then divided into "high" and "low" concentration groups based on whether or not each individual drug within the subset had a higher or lower concentration than the median for that drug. If the drug concentration equaled the median, then it was assigned to the "low" concentration group. If multiple different drugs of a single drug class subset (for example, if the opioid subset included methadone, hydrocodone, and oxycodone) were identified in the postmortem analysis of a particular case, that case would be classified as "high" concentration within its subset if at least one of the co-detected drugs was above the median. Given that drug metabolites are often pharmacologically active, for the purpose of this analysis, the various metabolites were treated independent from the parent drug.

To test the hypothesis that the concentration of morphine detected in a given case would be inversely proportional to the concentration of co-detected central nervous system depressant drugs, statistical comparison was made between morphine concentrations associated with the "high" versus "low" concentrations of a particular drug class subset. Probability testing included both a Mann–Whitney *U* (MW) test (http://elegans.swmed.edu~leon/stats/utest.html) and a one-tailed *t*-test (TT) with Microsoft Excel 2000 (Microsoft Corporation, Redmond, WA). Given the reliance on median value for the bulk of this analysis, MW calculations are probably the more reliable determination. Significance is considered $p < 0.05$.

Other factors were also evaluated and include age, height, and weight of each decedent. Also, to determine the impact of cocaine intoxication, each subset of cases was evaluated with and without the inclusion of cocaine and/or its major metabolite benzoylecgonine (BE) (abbreviated "cocaine/BE")-positive cases.

TABLE 1—*Drug class subsets. Listing of drugs included in each subset of drugs co-detected with opioids. Ethanol and cocaine are not shown.*

Co-Detected Drugs in Each Drug Class Subset			
Opioid Subset	Benzodiazepine Subset	Antidepressant Subset	"Other" Subset
Hydrocodone	Chlordiazepoxide	Citalopram	D-9 THC
Hydromorphone	Diazepam	Sertraline	11H D-9 THC
Oxycodone	Nordiazepam	Amitriptyline	D-9 Carboxy THC
Oxymorphone	Oxazepam	Nortriptyline	Olanzapine
Fentanyl	Lorazepam	Bupropion	Quetiapine
Norfentanyl	Temazepam	Fluoxetine	Chlorpromazine
Methadone	Alprazolam	Norfluoxetine	Mephobarbital
EDDP	Clonazepam	Doxepin	Phenobarbital
Norpropoxyphene	7-amino	Nordoxepin	Butalbital
Tramadol	Clonazepam	Fluvoxamine	Amphetamine
		Trazodone	Diltiazem
		Paroxetine	Metoprolol
			Cyclobenzaprine
			Carisprodol
			Diphenhydramine
			Hydroxyzine
			Trimethobenzamide
			Doxylamine
			Dextromthorphan
			Quinine
			Meprobamate

Results

In 2006, 13,422 cases were reported for investigation to the Commonwealth of Massachusetts OCME. Autopsy or external examination with or without toxicologic analysis was performed on 4396 of these cases. Of the 210 cases having autopsy reports that included the keyword "opiate" in the cause of death, 49 were eliminated from further evaluation. Reasons for elimination included: no detection of an opiate in available specimens (26 cases), availability of only matrices unacceptable for study inclusion including urine, decompositional fluid, or tissue samples (10 cases), fetal demise (1 case), manners of death that were not accidental or undetermined (4 cases), and insufficient information (8 cases). Ultimately, of the 210 cases in which the cause of death was found to be at least partially because of acute opiate intoxication, 161 (76.6%) cases were included for analysis.

Of the cases analyzed, 72.0% were men ($n = 116$) and 28.0% were women ($n = 45$), a ratio of 2.6. The age of the population ranged from 18 years to 79 years, with an average of 38.8 years. The manner of death was deemed accidental in 156 cases (96.9%) and undetermined in five cases (3.1%). Drug paraphernalia was documented at the location of death in 58 cases (36.0%), reportedly absent in 28 cases (17.4%) and was not confirmed present or absent by investigative reports in 75 cases (46.6%).

Quantitative analyses were performed using heart blood in 87 cases, femoral blood in 49 cases, a combination of heart and

femoral blood in eight cases, an unspecified blood source in seven cases, antemortem blood in seven cases, and pooled cavity blood, "peripheral blood," and subclavian blood in one case each.

Morphine was present in 159 cases (98.8%) and ranged from less than the limit of quantitation (20 ng/mL) to greater than the limit of quantitation (2000 ng/mL) with a mean concentration of 261.1 ng/mL and a median concentration of 168 ng/mL. In two cases where morphine was not identified, codeine was detected; no further analysis was performed on these cases. Morphine concentrations in cases that quantitated femoral blood (median = 124 ng/mL; mean = 202.9 ng/mL) versus cases that quantitated heart blood (median = 220 ng/mL; mean = 286.8 ng/mL) were significantly different (MW = 0.0004; TT = 0.0558).

There was no significant difference in the body weights of the decedents that corresponded to morphine concentrations equal to or less than the median morphine value (median = 78.9 kg; mean = 83.9 kg) versus those greater than the median morphine concentration (median = 83.9 kg; mean = 86.2 kg) (MW 0.2973; TT = 0.2416). The heroin metabolite, 6-monacetylmorphine, was detected in 11 cases with a mean concentration 17.8 ng/mL and a median of 16 ng/mL. Codeine was identified in 21 total cases (12.9%) with a mean concentration of 65.8 ng/mL and a median of 26 ng/mL.

Opioid Subset

Opioids were present in 30 cases (18.6%) (Table 2). The following opioids were present: hydrocodone, hydromorphone, oxycodone, oxymorphone, fentanyl, norfentanyl, methadone, EDDP, norpropoxyphene, and tramadol. Morphine concentrations ranged from <20 ng/mL to 1071 ng/mL with a mean of 206.7 ng/mL and a median of 124 ng/mL. In nine cases, cocaine/BE was also detected. There was no significant difference in morphine concentrations identified in samples with detectable postmortem concentrations of opioids equal to or less than the median value ("low" concentration) of the specific opioid (median = 155.5 ng/mL; mean = 243.4 ng/mL) versus those greater ("high" concentration) than the median (median = 103 ng/mL; mean = 172.3 ng/mL) (MW = 0.2161; TT = 0.2093) (Table 3).

Ethanol Subset

Ethanol was present in 50 cases (31.1%) and had mean and median concentrations of 0.11 gm% (Table 4). Morphine concentrations ranged from 30 ng/mL to >2000 ng/mL with a mean of 292.7 ng/mL and a median of 202 ng/mL. In 23 cases, cocaine/BE was also detected. There was no significant difference in morphine concentrations identified in samples with detectable postmortem concentrations of ethanol equal to or less than the median value of ethanol (median = 176.5 ng/mL; mean = 241.4 ng/mL) versus those greater than the median (median = 220; mean = 361.1 ng/mL) (MW = 0.1562; TT = 0.1313) (Table 3).

Benzodiazepine Subset

Benzodiazepines were present in 35 cases (21.7%) (Table 5). The following benzodiazepines were detected: chlordiazepoxide, diazepam, nordiazepam, oxazepam, lorazepam, temazepam, alprazolam, clonazepam, and 7-amino clonazepam. Morphine concentrations ranged from <20 ng/mL to 633 ng/mL with a mean of 215.7 ng/mL and a median of 143.5 ng/mL. In 14 cases, cocaine/BE was also detected. There was no significant difference in

TABLE 2—*Opioid-positive cases, including number of cases (*n*), median and mean opioid concentrations, and range of morphine concentrations corresponding to that particular opioid.*

Opioid Subset			
Co-Detected Opioids	*n*	[Opioid] Median/ Mean (ng/mL)	[Morphine] Median/ Mean (range) (ng/mL)
Methadone	16	153.5/238.4	126.5/199.4 (32–573)
EDDP	11	77.5/51.4	116.5/208.6 (32–573)
Fentanyl	6	21/18.9	80/74.5 (<20–124)
Oxycodone	5	58/72.8	181/216.4 (35–573)
Oxymorphone	2	88/54	64, 229
Norfentanyl	2	3.3/3.3	<20, 49
Norpropoxyphene	2	0.61/0.61	185, 412
Hydrocodone	1	58/58	1071
Tramadol	1	486/486	288

TABLE 3—*Summary table showing the number of cases detected in each drug class subset (*n*), mean and median morphine concentrations for the [High] and [Low] groups of each subset, and the corresponding* p *value (Mann–Whitney* U*).*

Morphine Concentrations Corresponding to [High] and [Low] Drug Class Subsets				
Drug Class Subsets	*n*	Morphine Conc. of [High] Drug Class Subset Median/Mean (ng/mL)	Morphine Conc. of [Low] Drug Class Subset Median/Mean (ng/mL)	*p*
Opioids	46	103/172.3	155.5/243.4	0.216
Ethanol	50	220/361.1	176.5/241.4	0.156
Benzodiazepines	59	165/243.8	135/184.1	0.186
Antidepressants	53	321.5/514.6	120/202.9	0.032

TABLE 4—*Ethanol-positive cases, including number of cases (*n*), median and mean ethanol concentrations, and range of corresponding morphine concentrations.*

Ethanol Subset			
Co-Detected Ethanol	*n*	[Ethanol] Median/ Mean (gm%)	[Morphine] Median/Mean (range) (ng/mL)
Ethanol	50	0.11/0.11	202/292.7 (30–>2000)

TABLE 5—*Benzodiazepine-positive cases, including number of cases (*n*), median and mean benzodiazepine concentrations, and range of morphine concentrations corresponding to that particular drug.*

Benzodiazepine Subset			
Co-Detected Benzodiazepine	*n*	[Benzodiazepine] Median/Mean (ng/mL)	[Morphine] Median/ Mean (range) (ng/mL)
Nordiazepam	19	200/352.8	130/200.4 (35–745)
Alprazolam	12	35/76.2	159/170.5 (<20–412)
Diazepam	10	160/337.6	129.5/186.4 (111–633)
Chlordiazepoxide	5	2000/2732	185/304 (76–745)
7-amino Clonazepam	5	71/112.2	170/224.8 (64–573)
Oxazepam	3	61/75.7	129/147.7 (122–192)
Temazepam	2	86/86	122, 129
Clonazepam	2	8.8/8.8	64, 88
Lorazepam	1	11/11	168

morphine concentrations identified in samples with detectable postmortem concentrations of benzodiazepines equal to or less than the median value of the specific benzodiazepine (median = 135 ng/mL; mean = 184.1 ng/mL) versus those greater than the median (median = 165 ng/mL; mean = 243.8 ng/mL) (MW = 0.1311; TT = 0.1859) (Table 3).

Antidepressant Subset

Antidepressants were present in 29 cases (18.0%) (Table 6). The following antidepressants were detected: citalopram, sertraline, amitriptyline, nortriptyline, bupropion, fluoxetine, norfluoxetine, doxepin, nordoxepin, fluvoxamine, trazodone, and paroxetine. Morphine concentrations ranged from 22 ng/mL to >2000 ng/mL with a mean of 358.8 ng/mL and a median of 221.5 ng/mL. In 13 cases, cocaine/BE was also detected.

There was a significant difference (MW = 0.0108; TT = 0.0318) in morphine concentrations identified in samples with detectable postmortem concentrations of antidepressants equal to or less than the median value of the specific antidepressant (median = 120 ng/mL; mean = 202.9 ng/mL) versus those greater than the median (median = 321.5 ng/mL; mean = 514.6 ng/mL) (Table 3). This difference in morphine between the "high" and "low" antidepressant concentration groups remained even after excluding the trazodone- and fluvoxamine-positive cases (MW = 0.0064; TT = 0.0285).

Within the antidepressant subset, there was not a significant difference between the ages (MW = 0.4652; TT = 0.4724) or genders (MW = 0.3234; TT = 0.3022) associated with the antidepressant concentrations equal to or less than the median values (median = 45 years, mean = 42.5 years, number of men = 9, number of women = 6) versus those greater than the median antidepressant values (median = 46.5 years, mean = 42.8 years, number of men = 7, number of women = 7). There was also not a significant difference between the morphine concentrations in cases that analyzed heart blood (median = 108 ng/mL; mean = 508.7 ng/mL) versus those that analyzed from femoral blood (median = 229 ng/mL; mean = 284 ng/mL) (MW = 0.2503; TT = 0.2312). There was a significant difference between the body weights associated with the antidepressant concentrations equal to or less than the median values (median = 83.9 kg; mean = 74.8 kg) versus those greater than the median values (median = 94.3 kg; mean = 99.3 kg) (MW = 0.0334; TT = 0.0214).

TABLE 6—*Antidepressant-positive cases, including number of cases (*n*), median and mean antidepressant concentrations, and range of morphine concentrations corresponding to that particular drug.*

Antidepressant Subset			
Co-Detected Antidepressant	*n*	[Antidepressant] Median/Mean (ng/mL)	[Morphine] Median/ Mean (range) (ng/mL)
Nortriptyline	8	330/557.1	689.5/822 (108–>2000)
Norfluoxetine	8	450/1808.8	111/121.6 (46–297)
Amitriptyline	6	395/475	379/786.4 (108–>2000)
Citalopram	6	405/562.3	258/320.1 (22–1000)
Fluoxetine	6	365/391.7	130/100.6 (46–165)
Doxepin	4	605/837.5	165/190 (46–384)
Nordoxepin	4	340/377.5	165/190 (46–384)
Paroxetine	4	230/375	252/402 (104–>1000)
Bupropion	3	140/170.7	214/465 (110–1071)
Sertraline	2	2100/2100	773, 1071
Trazodone*	1	0.61/0.61	104
Fluvoxamine	1	2800/2800	170

*Concentrations of trazodone reported in mcg/mL.

Other Subset

A heterogenous group of CNS depressant drugs that could not be classified in the other subsets of this study were included in the "other" category and account for 29 of the cases (17.8%) (Table 7). Meaningful statistical analysis of this group was precluded because of the low number (often only one) of each individual identified compound and the wide diversity of drug mechanisms. Morphine concentrations ranged from <20 ng/mL to 1071 ng/mL with a mean of 240.3 ng/mL and a median of 155 ng/mL. In 13 cases, cocaine/BE was also detected.

Cocaine-Positive vs. Cocaine-Negative Groups

Cocaine and/or its major metabolite, BE, were present in 71 cases (44.1%) and absent in 90 cases (55.9%). The average age of decedents with detection of morphine and cocaine/BE is 37.3 years (median = 39 years) while the average age for those without detectable cocaine/BE is 40 years (median = 43 years) (MW = 0.0665; TT = 0.0692). Morphine concentrations in the cocaine/BE-positive group ranged from less than (20 ng/mL) to greater than (1000 ng/mL) the limits of quantitation, with a median morphine concentration of 129 ng/mL and a mean morphine concentration of 198.8 ng/mL. Morphine concentrations in the cocaine/BE-negative group also spanned the limits of quantitation with a median morphine concentration of 197 ng/mL and a mean of 305.6 ng/mL. The morphine concentrations between the cocaine/BE-positive and negative groups did vary significantly (MW = 0.0021; TT = 0.0057). There was no significant difference (MW = 0.1395; TT = 0.2385) between the morphine concentrations corresponding to cocaine/BE-negative cases and cases positive for BE (but not the parent drug).

When cases with cocaine and/or its metabolites were excluded from analysis, there was no significant difference between morphine concentrations in cases with "high" versus "low" concentrations of co-detected opioids (MW = 0.4387, TT = 0.2809),

TABLE 7—*Number of cases present for each "other" drug detected (*n*), median and mean "other" drug concentrations, and range of morphine concentrations corresponding to that particular drug.*

Other Subset			
Co-Detected "Other" Drug	*n*	[Other] Median/ Mean (ng/mL)	[Morphine] Median/ Mean (range) (ng/mL)
D-9 Carboxy THC	13	12/16	138/231.9 (22–1000)
D-9 THC	6	3.7/9.5	89.5/111 (22–278)
Diphenhydramine	4	460/702.5	58/50.7 (30–168)
Olanzapine	2	84/84	41, 111
11H D-9 THC	2	2/2	41, 199
Chlorpromazine	1	140/140	207
Mephobarbital	1	0.79/0.79	49
Phenobarbital	1	1.8/1.8	49
Butalbital	1	3.2/3.2	202
Amphetamine	1	86/86	168
Diltiazem	1	890/890	1071
Metoprolol	1	180/180	773
Cyclobenzaprine	1	700/700	1071
Carisprodol	1	6/6	392
Quetiapine	1	1000/1000	207
Hydroxyzine	1	310/310	392
Trimethobenzamide	1	0.95/0.95	<20
Doxylamine	1	64/64	58
Dextromethorphan	1	150/150	258
Quinine	1	2.8/2.8	185
Meprobamate	1	13/13	392

benzodiazepines (MW = 0.1121, TT = 0.1601), and ethanol (MW = 0.0955, TT = 0.4516). While significance is not established, there does appear to be a trend toward a difference in morphine concentrations in the "high" and "low" concentration groups of antidepressants (MW = 0.0704, TT = 0.1124).

Discussion

The frequency of fatal heroin overdoses is increasing throughout Massachusetts and nationally (2,3). This study investigates trends in drug concentrations of opiates and other psychotropic drugs in the setting of fatal polydrug overdose. Similar to other heroin overdose studies, men were over-represented (72.3%) in our population (4,5,10). The average age of victims of fatal heroin overdose typically falls between the late twenties and early thirties (5). However, similar to the average age of 38.9 years in our population, a study by Oppenheimer et al. (11) has reported an average age of victims of fatal heroin overdose as 38 years. This further supports that most fatal heroin overdoses occur in older, experienced heroin users rather than younger, and inexperienced users (5).

There was a significant difference between the morphine concentrations quantified from heart blood versus those quantified from femoral blood. While this observation is between cases rather than within a single case, the finding is still not unexpected. Morphine is primarily metabolized and distributed throughout the organs of the torso, which after death redistribute the drug by diffusion into central (heart) blood, causing an increase in postmortem concentrations. The femoral vessels are less susceptible to postmortem redistribution because the only theorized source of concentrated drugs is the surrounding muscle tissue. Therefore, femoral blood will typically have a lower concentration (that purportedly better mirrors antemortem concentration) than heart blood (12).

Contrary to the hypothesis that the concentration of morphine would be inversely proportional to the concentration of co-detected central nervous system depressant drugs in the setting of polydrug overdoses, this study does not identify distinct trends in the opioid, ethanol, and benzodiazepine-positive subsets. Specifically, it was expected that the concentration of morphine would be lower in the "high" concentration cases of a particular central nervous system depressant drug class subset, and higher in the "low" concentration cases. This expectation was based on the belief that given their similar mechanisms and effects, central nervous system depressant drugs would react synergistically in a concentration-dependent manner with morphine and that the cumulative result would be an overdose fatality. Particularly, the presence of opioids was believed to result in a fatal overdose with a lesser concentration of morphine, because each targets the same μ receptors with resulting respiratory depression (13). Benzodiazepines (and barbiturates) produce effects by targeting the GABAA receptor complex, which also decreases the respiratory rate. Although the respiratory inhibition of benzodiazepines alone is not as great as the effect of opioids alone, significant respiratory depression can occur when both the μ and GABAA receptors are stimulated simultaneously (13). Alcohol also acts through the GABAA receptor but binds through an alternate site than benzodiazepines. Like benzodiazepines, alcohol has respiratory depressive effects that are compounded with the addition of opiates (13). However, in this study, there were no significant findings among these subsets that demonstrate evidence of quantitative synergism with morphine.

It is possible that trends were not uncovered in the opioid, ethanol, and benzodiazepine subsets because of a lack of statistical power. In small populations, the addition of extra values to the set can of course have a considerable effect on the mean, and to a lesser degree on the median. The use of a nonparametric statistical test (Mann–Whitney *U*), which relies upon median values, is useful in this regard.

Concomitant use of heroin and antidepressants were detected in 17.8% of the total cases in this study. This figure is much higher than those discovered in an earlier study by Darke et al. (14) where antidepressants were detected in only 7% of fatal heroin overdoses. A separate study by Darke and Ross (15) investigated the habits of current heroin users, citing depression as the most common reason (42%) heroin users ingested antidepressants. Other heroin users reported that antidepressants were taken for intoxication (12%), sleep, anxiety, management of heroin withdrawal, and to reduce benzodiazepine use (15).

Within the antidepressant subset in this study, a direct relationship was found between the morphine values corresponding to an antidepressant concentration equal to or below the median antidepressant concentration versus the morphine concentrations corresponding to antidepressant concentrations greater than the median antidepressant concentration. In other words, morphine concentrations were higher in the "high" concentration group of the antidepressant drug class subset and lower in the "low" concentration group. These findings may suggest that the lethal dosage of heroin is increased through heavier concomitant antidepressant use.

Earlier studies have linked tricyclic antidepressants to potentiation of the effects of opiates because of the interaction with serotonin and other neurotransmitters that activate the endogenous opioid system and increase endorphin levels (16). Selective serotonin reuptake inhibitors (SSRI) also have toxic effects when used with heroin although to a lesser degree than tricyclic antidepressants (17).

A study by Fialip et al. (18) with mice found that the effects of morphine on the central nervous system were enhanced by the administration of clomipramine, a tricyclic antidepressant, but only if the clomipramine was administered simultaneously or shortly before the morphine. Alternatively, the same study found that chronic administration of clomipramine inhibited the effects of morphine. The authors concluded that the inhibitory effects of chronic clomipramine administration on morphine analgesia are time-dependent for a given dose and suggested some type of "gradual neurobiochemical changes first suppressing the potentiating effect and then reversing it." They proposed that their findings might be due to a specific loss of opiate receptor binding sites in the cerebral cortex after chronic tricyclic antidepressant administration. After 21 days of administration of a tricyclic antidepressant (desimipramine) in rats, Reisine and Soubrie (19) showed a significant decrease in sensitivity to opiates because of the loss of the specific opiate receptors in the cerebral cortex but not the corpus striatum or the hippocampus.

It was beyond the scope of this study to determine the length of antidepressant use prior to fatal overdose. Regardless, the finding that chronic tricyclic antidepressant administration decreases the quantity of opiate receptors in the cerebral cortex thus inhibiting the euphoric effects of morphine by Reisine and Soubrie (19) serves as a possible explanation for the direct relationship seen in our study between morphine and antidepressant concentrations. Heroin users on antidepressant therapy may increase their dosages in an unsuccessful attempt to gain heightened euphoria, when in actuality their opiate receptors in the cortex are subsensitive. Meanwhile, the opiate receptors of the corpus striatum and hippocampus, which seem relatively unaffected by chronic antidepressant use, become overwhelmed by the increase in opiate dosage. Although the hippocampus is not typically considered to have a direct role in respiration, the hippocampus does play a part in the rate and timing

of breathing. A study by Harper et al. (20) has shown that the theta waves produced by the hippocampus increase during inspiratory activity following a pause in the normal breathing pattern. This finding suggests that the hippocampus plays a critical role in mediating changes in the breathing pattern by producing the "drive" to breathe. Activation of the opiate receptors of the hippocampus can apparently inhibit the driving force to restart inspiratory activity following an anoxic period, thus leading to fatal respiratory depression.

Selective serotonin reuptake inhibitors were developed as a safer alternative to tricyclic antidepressants and their efficacy remains even when combined with morphine (21–23). One study found that the effects of morphine were unchanged in rats during the chronic administration of fluvoxamine (22). Another study found that fluoxetine actually increased the analgesic effects of morphine in healthy humans (23). Given the conflicting available data, as well as the results of this work, more studies are essential to determine whether antidepressants may inhibit the pleasurable effects of morphine in the cerebral cortex, thus requiring users of both to increase morphine doses. Alternatively, it could be postulated that antidepressant use may actually protect against respiratory depression, thus allowing the heroin user to increase morphine doses.

A striking finding of this study was the large number of cases of fatal overdose involving cocaine and heroin. Both morphine and cocaine (or BE) were present in nearly half (44.1%) of the cases in this study; a combination often termed a "speedball." A study by Foltin and Fischman (24) found that drug users prefer to use speedballs because the combination reduces the severity of the "crash" after cocaine use, takes the "edge" off of cocaine, and prolongs the "high" of the drugs, although the combined effects were not determined to have a significantly greater effect than each drug alone. However, a more recent study found that the combination of heroin and cocaine resulted in lower concentrations of morphine causing a lethal effect (25). These results correspond to the findings of this study, namely that significantly lower morphine concentrations are found in cocaine-positive cases and vice versa. Polettini et al. (25) speculated that these findings are the result of heroin and cocaine interfering with the metabolism of one another as they are metabolized through the same liver carboxylesterases. Further studies are needed to verify this hypothesis, as well as to further characterize the mechanism of fatality from this drug combination.

There are limitations to this study. It is retrospective; therefore, particular valuable information may not have been collected. The small number of cases in some of the subsets may have precluded the ability to detect important statistical differences. The use of keyword searching of causes of death may have omitted cases that fit the inclusion criteria. The route of administration was not determinable in every case, thus precluding insight into current usage trends or speculation as to how the rate of absorption may have played a role in the lethal event. Studies have shown that smoking rather than intravenously injecting heroin may be increasing in popularity and that different routes of administration cause a drug to achieve different levels of effects (5). Toxicological analyses in this study were performed on a variety of biological matrices samples, which affects the overall uniformity of the results. The use of antemortem samples ($n = 7$), although ultimately allowed in this study and not particularly outliers, may skew results. Even though the antemortem samples were not exposed to postmortem redistribution, they were included because they represented samples taken at a hospital immediately prior to the expiration of the patient and therefore, may be the most accurate representation of drug concentrations at the time of death. A similar prospective study that accounted for these variables may be of value.

Conclusion

Drug abuse is prevalent nationwide with heroin-related fatalities on the rise particularly. In the fatal heroin overdose population, polydrug fatalities where opiates are combined with other central nervous system agents are very common. This study reveals that there is no detectable relationship between postmortem morphine concentrations and the concentrations of other central nervous system depressants, namely opioids, benzodiazepines, and ethanol. Surprisingly, there is a directly proportional relationship between postmortem morphine and antidepressant concentrations. This relationship may be related to the decrease of opiate receptors in the cerebral cortex after chronic antidepressant treatment. While certainly further studies are needed, these findings may prove useful in understanding addiction and polydrug overdoses, as well as may be important in the treatment of depression in heroin addicts.

References

1. Boston Public Health Commission. The health of Boston 2008. Boston, MA: Boston Public Health Commission, Research Office, 2008.
2. Crane E, Novak S. Opiate-related drug misuse deaths in six states: 2003. The New DAWN Report 2006;19:1–4. Available at: http://dawninfo.samhsa.gov/files/tndr/2006-06/tndr06opiatemisuse.htm. Accessed March 21, 2010.
3. National Drug Intelligence Center. Massachusetts drug threat assessment (Document ID: 2001-S0377MA-001); April 2001.
4. Warner-Smith M, Darke S, Lynskey M, Hall W. Heroin overdose: causes and consequences. Addiction 2001;96:1113–25.
5. Darke S, Zador D. Fatal heroin 'overdose': a review. Addiction 1996;91(12):1765–72.
6. Darke S, Hall W. Levels and correlates of polydrug use among heroin users and regular amphetamine users. Drug Alcohol Depend 1995;39:231–5.
7. Darke S, Hall W. Heroin overdose: research and evidence-based intervention. J Urban Health 2003;80(2):189–200.
8. Mather LE. Opioid pharmokinetics in relation to their effects. Anaesth Intensive Care 1987;15:15–22.
9. Hull MJ, Juhascik M, Mazur F, Flomenbaum MA, Behonick GS. Fatalities associated with fentanyl and co-administered cocaine and opiates. J Forensic Sci 2007;52:1383–8.
10. Fugalstad A, Ahlner J, Brandt L, Ceder G, Eksborg S, Rajs J, et al. Use of morphine and 6-monoacetylmorphine in blood for the evaluation of possible risk factors for sudden death in 192 heroin users. Addiction 2003;98:463–70.
11. Oppenheimer E, Tobutt C, Taylor C, Andrew T. Death and survival in a cohort of heroin addicts from London clinics: a 22-year follow-up study. Addiction 1994;89(10):1299–308.
12. Cook DS, Braithwaite RA, Hale KA. Estimating antemortem drug concentrations from postmortem blood samples: the influence of postmortem redistribution. J Clin Pathol 2000;53:282–5.
13. White JM, Irvine RJ. Mechanisms of fatal opioid overdose. Addiction 1999;94(7):961–72.
14. Darke S, Ross J, Zador D, Sunjic S. Heroin-related deaths in New South Wales, Australia, 1992-1996. Drug Alcohol Depend 2000;60(2):141–50.
15. Darke S, Ross J. The use of antidepressants among injecting drug users in Sydney, Australia. Addiction 2000;95(3):407–17.
16. Hepburn S, Harden J, Grieve JHK, Hiscox J. Deliberate misuse of tricyclic antidepressants by intravenous drug users—case studies and report. Scott Med J 2005;50(3):131–3.
17. Battersby MW, O'Mahoney JJ, Bethwick AR, Hunt JL. Antidepressant deaths by overdose. Aust N Z J Psychiatry 1996;30:223–8.
18. Fialip J, Marty H, Makambila M-C, Civiale M-A, Eschalier A. Pharmacokinetic patterns of repeated administration of antidepressants in animals. II. Their relevance in study of the influence of clomipramine on morphine analgesia in mice. J Pharmacol Exp Ther 1988;248(2):747–51.
19. Reisine T, Soubrie P. Loss of rat cerebral cortical opiate receptors following chronic desipramine treatment. Eur J Pharmacol 1982;77(1):39–44.
20. Harper RM, Poe GR, Rector DM, Kristensen MP. Relationships between hippocampal activity and breathing patterns. Neurosci Biobehav Rev 1998;22(2):233–6.

21. Krupitsky EM, Burakov AM, Didenko TY, Romanova TN, Grinenko NI, Slavina TY, et al. Effects of citalopram treatment of protracted withdrawal (syndrome of anhedonia) in patients with heroin addiction. Addict Disord Their Treat 2002;1(1):29–33.
22. Gutierrez M, Ortega-Alvaro A, Gibert-Rahola J, Mico JA. Interactions of acute morphine with chronic imipramine and fluvoxamine treatment on the antinoceptive effect in arthritic rats. Neurosci Lett 2003;352:37–40.
23. Erjavec MK, Coda BA, Nguyen Q, Donaldson G, Risler L, Shen DD. Morphine–fluoxetine interactions in healthy volunteers: analgesia and side effects. J Clin Pharmacol 2000;40:1286–95.
24. Foltin RW, Fischman MW. The cardiovascular and subjective effects of intravenous cocaine and morphine combinations in humans. J Pharmacol Exp Ther 1992;261(2):623–32.
25. Polettini A, Poloni V, Groppi A, Stramesi C, Vignali C, Politi L, et al. The role of cocaine in heroin-related deaths: hypothesis on the interaction between heroin and cocaine. Forensic Sci Int 2005;153:23–8.

Additional information—reprints not available from author:
William J. Minett, M.S.
Office of the Chief Medical Examiner
720 Albany Street
Boston, MA 02118
E-mail: minettw@hotmail.com

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01429.x
Available online at: interscience.wiley.com

TECHNICAL NOTE

PHYSICAL ANTHROPOLOGY

Can Pelin,[1] *M.D., Ph.D.; Ragıba Zağyapan,*[1] *Ph.D.; Canan Yazıcı,*[1] *M.D., Ph.D.; and Ayla Kürkçüoğlu,*[1] *M.D., Ph.D.*

Body Height Estimation from Head and Face Dimensions: A Different Method*

ABSTRACT: As there are cases brought for forensic examination where only the craniofacial region is available, estimation of stature from craniofacial dimensions is without doubt important in forensic cases. The study presented here attempts to estimate stature from craniofacial dimensions in the Turkish population. In the second phase of the study, the correlations between craniofacial dimensions and stature were also evaluated according to different head and face types. All measurements were taken from 286 healthy males with a mean age of 22.71 ± 4.86 years. The sample was then reclassified according to different head and face indexes. For the whole sample, correlation coefficients were low, changing only between 0.012 and 0.229. Thus, no significant increase in correlation coefficients was observed after the samples had been reevaluated according to different head and face types. As a conclusion, craniofacial dimensions are not good predictors for body height for the Turkish population.

KEYWORDS: forensic science, forensic anthropology, stature estimation, head dimensions, face dimensions, identification

It is well known that the accurate investigation of stature helps to establish an individual's identity in medicolegal investigations involving skeletal remains. Being one of the criteria of personal identification, stature helps in narrowing down the investigation process, and thus provides useful clues for the investigating institution. The anatomical and mathematical methods are the two main techniques for the estimation of living stature (1,2). The anatomical method involves the direct reconstruction of stature by measuring and adding together the lengths or heights of all the skeletal elements from the skull to the foot and figuring a correction factor for soft tissues (1,2). The method is believed to provide the best approximation when it is applicable (1,3,4). However, it is not always possible to obtain the whole skeleton. The mathematical method is based upon the proportion of certain bones to living stature or cadaver length because proportions do not alter with age after skeletal maturity had been fully completed (5). However, body height does decrease with age, and age-correcting factors are used during stature estimations for older people.

Currently, the mathematical method is used worldwide both in forensic and in biological anthropology. This method was first used by Rolet in 1889 and later reevaluated by Manouvrier in 1892 and Pearson in 1899 (6,7). Pearson was the first person to introduce regression formula for estimating living stature from long limb bones. Since then many reports have appeared suggesting the number of formulae devised to compute stature from long bones in various countries. In 1978, Krogman (8) summarized the work of various anthropologists who had calculated stature from long bones. Numerous researchers continue to study this subject (9–13).

[1]Departments of Anatomy and Biostatistics, Faculty of Medicine, Baskent University, Baglıca-Ankara, Turkey.

*Presented at the XII National Congress of Anatomy, October 29–November 1, 2008, in Mersin, Turkey.

Received 25 Mar. 2009; and in revised form 3 July 2009; accepted 23 July 2009.

Long bones, especially those of lower limbs are considered the best indicator of living stature. However, intact main long bones of limbs are not always available. Several regression equations using the measurements from other postcranial bones or bone segments have been calculated to estimate body height in the absence of intact long bones. Other than long bones, hand and foot dimensions are the most commonly used anthropometric measurements for stature estimation (5,14,15). Even measurements from isolated bones of the hand or foot such as the metacarpal (16–19) or tarsal bones have been evaluated as independent variables for the estimation of body height (20–22). Not only the absolute anthropometric dimensions but also measurements such as arm span or sole length have also been used for estimating body height (14,23,24).

Measurements from the members of the axial skeleton for stature determination are not so common. It has been reported that sacral dimensions and measurements from sacral segments can be used with moderate accuracy to predict stature (25,26). Only a few studies have reported on reconstructing stature from head and face dimensions (27–29). However, there are cases brought for forensic examination where only the craniofacial region is available. Additionally, in the field of archeology, more skulls are commonly found intact during excavations compared to the other parts of the human skeleton. Chiba and Terazawa reported a significant correlation between stature and skull dimensions (27). In another study on lateral cephalographs by Patil and Mody, regression equations based on skull length gave reliable estimations (28). However, in their study on South African skulls, Ryan and Bidmos reported only a moderate correlation between skull measurements and skeletal height (20,29). The last study on cranial dimensions related with body height using a large sample (996 male Indians) was the one by Krishan. Correlation coefficients found by Krishan were higher than all other studies on this subject (30). On the other hand, Sarangi et al. (31) reported no significant findings between body height and skull dimensions.

It is accepted that studies of stature estimation based on an isolated body part should be population specific. Research on the prediction of body height from head and face dimension for the Turkish population had not been reported to date. The main purpose of this study was to evaluate the correlations between body height and craniofacial dimensions in the Turkish population and to calculate regression equations for the estimation of stature to establish a reliable identification method in the forensic field.

It should be kept in mind that there can be several head or face types in a population. In studies to date, correlations between craniofacial dimensions and stature have not been evaluated in relation to face and head types. In the second step of this study, these correlations were evaluated using different cephalic and facial indexes for the purpose of calculating cranial- or facial-type-specific regression equations to achieve more reliable predictions.

Materials and Methods

The study was conducted on 286 randomly selected healthy males. None of the subjects had had a serious craniofacial injury or a surgical procedure. The mean age and standard deviation of the sample were 22.71 ± 4.86 years (min = 18, max = 45). To eliminate the effects of age on stature, anthropometric measurements were not taken from individuals over 45 years of age. All the subjects were from various socioeconomic backgrounds and cities of Anatolia, but living in the central districts of Ankara at the time of the study. In addition to body height, the below-mentioned nine anthropometric measurements were taken from each subject according to Olivier's methods (32).

Maximum head length (MHL): The distance between glabella and opisthocranion.

Maximum head breadth (MHB): The distance between the most lateral points on the parietal bones (euryon) on each side of the head.

Minimum frontal diameter (MFD): The least breadth of the forehead between the two frontotemporal points on the temporal ridges.

Horizontal circumference of head (HC): It is measured from glabella to glabella with the measuring type passing over opistocranion.

Maximum head height (MHH): The distance between the vertex of the head to the upper border of tragus (tragion).

Bizygomatic breadth (BZB): The maximum distance between the most lateral points on the zygomatic arcs.

Bigonial diameter (BGD): The maximum breadth of the lower jaw between the two gonion points on the angles of mandible.

Morphological facial length (FL): The distance between the nasal root (nasion) and the lowest point on the lower border of mandible in the midsagittal plane (gnathion).

Morphological superior facial length (SFL): The distance between the nasal root (nasion) and the gum between the upper central teeth (prosthion) in the midsagittal plane.

Stature was measured with the subjects standing in bare feet with their back to an anthropometer. The subject's head was adjusted to the Frankfurt horizontal plane, and then the head was tilted slightly upward by applying gentle force to the mastoid process and zygomatic bones (33).

All measurements were taken by the same researcher, at approximately the same time of day, and recorded to the nearest millimeter. A Martin-type anthropometer set was used for the measurements. Descriptive statistics for the variables are shown in Table 1.

TABLE 1—*Descriptive statistics for the variables (*n = *286).*

Variable	Mean ± SD	Minimum	Maximum
Age	22.71 ± 4.86	18	45
Stature	1753.14 ± 61.70	1580	1911
Max. head length	188.84 ± 7.37	171.00	212.00
Max. head breadth	156.03 ± 6.89	134.00	181.00
Min. frontal diameter	117.24 ± 6.35	101.00	136.00
Head circumference	564.01 ± 15.14	517.00	617.00
Head height	110.50 ± 11.65	75.00	146.00
Bizygomatic breadth	142.91 ± 6.07	121.00	161.00
Bigonial diameter	103.61 ± 9.89	70.00	125.00
Morphological facial length	121.70 ± 7.16	105.00	145.00
Morphological superior facial length	72.51 ± 5.29	58.00	92.00

SD, standard deviation.

TABLE 2—*Cephalic indexes used in the study.*

100 × maximum head breadth/maximum head length	
Hyperdolicocephalic	Up to 70.9
Dolicocephalic	71.0–75.9
Mesocephalic	76.0–80.9
Brachycephalic	81.0–85.4
Hyperbrachycephalic	Above 855
Height–length index (100 × head height/head length)	
Chamaecephalic	Up to 57.6
Orthocephalic	57.7–62.5
Hypsicephalic	Above 62.6
The height–breadth index (100 × head height/head breadth)	
Tapeinocephalic	Up to 78.9
Metriocephalic	79.0–84.9
Acrocephalic	Above 85.0
The mean index of height (100 × head height/half the sum of breadth and length	
Low	Up to 67.9
Medium	68.0–72.9
High	Above 73.0

TABLE 3—*Facial indexes used in the study.*

100 × face height/bizygomatic breadth (BZB)	
Hypereuryprosope	Up to 78.9
Euryprosope (broad or low face)	79.0–83.9
Mesoprosope (medium face)	84.0–87.9
Leptoprosope (narrow or high face)	88.0–92.9
Hyperleptoprosope	Above 93.0
100 × superior facial length/BZB	
Euryene (broad face)	Up to 47.9
Mesene (medium face)	48.0–52.9
Leptene (narrow face)	Above 53.0
Zygomandibular index (100 × bigonial breadth/BZB)	
Narrow jaw	Up to 75.9
Medium jaw	76.0–77.9
Wide jaw	Above 78.0
Transverse cephalofacial index (100 × BZB/max. head breadth)	
Micropside (narrow)	Up to 89.9
Mesopside (medium)	90.0–92.9
Macropside (broad)	Above 93.0

For the later steps of the study, the sample was reclassified according to cephalic (Table 2) and facial (Table 3) indexes (32).

The correlations between stature and head/face dimensions were evaluated for the components of each index. Distributions of the variables were analyzed using the Shapiro–Wilk's normality test. Variables were distributed normally, so Pearson correlation coefficients were calculated. Statistical analyses were performed by SPSS 13.0 statistical software (Chicago, IL).

Results

When the correlations between stature and head/face dimensions were evaluated for the general population, MHL gave the highest correlation with body height (0.229). This was followed by morphological facial height (0.199) and head circumference (0.196). No correlation was found between MHB and MFD, and stature.

The sample was then divided into four subgroups according to head breadth/head length index. The calculations indicated that most of the individuals in the sample were brachycephalic. However, the percentages of hyperbrachiocephals and mesocephals were close to that of brachiocephalic individuals. Next, the correlations between the measurements from head and stature were determined separately for each subgroup and compared with those of the entire sample (Table 4). There was no significant increase in the correlation values specific to the subgroups. In fact, even some decreases were observed for some subgroups. As illustrated in Table 4, only in brachiocephalic individuals was a slight increase observed in the correlation between head length, head breadth, head circumference, and stature. However, no significant correlation was found between stature and head dimensions for the other head types.

When the sample was classified according to height/length index, in hypsicephalic subjects, an increase was observed in the correlation values for head length, head circumference, and head height. Similarly, a slight increase was observed in the correlations of head length and head height, and body height in orthocephalic individuals (Table 5).

When the sample was classified according to the mean index of height (100 × head height/half of the sum of breadth and length) only in the medium-type subjects did the correlation between head length–stature, head height–stature, and head circumference–stature increase (Table 6). However, most of the subjects in the sample had a low mean index of height value.

To evaluate the correlation between face dimensions and stature, the sample was first classified depending on facial index (100 × face height/BZB). The findings of the study indicated that most of the male subjects from Turkey had a medium-type face. This was followed by broad or low face type. Correlations between face dimensions and stature were then evaluated for each face type (Table 7). As shown in Table 7, only for individuals with a narrow or high face did the correlation between face dimensions and stature significantly increase for all variables. Correlations for other face types significantly decreased. Specifically, for individuals belonging to hypereuryprosope, mesoprosope, and hyperleptoprosope groups, it is almost impossible to estimate stature depending on face dimensions.

The sample was later classified according to the zygomandibular index (100 × BGD/BZB), an index representing the shape of the lower jaw with respect to the face (Table 8). When specific correlation values for the subgroups of zygomandibular index were evaluated, only a slight increase was observed in the correlation value of face height and stature in the narrow jaw group. For individuals with a wide jaw, the correlation between BZB and BGD, and body height increased. However, most of the subjects in the sample had a narrow jaw. Estimation of stature depending on face dimensions is almost impossible for individuals who have a medium jaw.

TABLE 5—*The correlations between the head dimensions and stature for the components of height/length index (100 × head height/head length).*

Variable	General Population (n = 286)	Height/Length Index (100 × Head Height/Head Length)		
		Chamaecephalic (n = 117)	Orthocephalic (n = 99)	Hypsicephalic (n = 70)
MHL	0.229**	0.123	0.295**	0.403**
MHB	0.012	0.027	−0.116	0.160
MFD	0.012	−0.31	0.019	0.084
HC	0.196**	0.144	0.195	0.329**
MHH	0.162**	0.149	0.249*	0.298*

*$p < 0.05$; **$p < 0.001$. MHL, maximum head length; MHB, maximum head breadth; MFD, minimum frontal diameter; HC, horizontal circumference; MHH, maximum head height.

Discussion

In addition to age, sex, and ethnical origin, estimation of body height is without doubt of importance for reliable identification in the forensic field. The most accurate estimations have been reached by formulae based on the long bones of limbs. However, when long bones were not available, measurements from other bones or body parts have been used to predict body height. Studies concerning the estimation of stature from head and face dimensions are not so common.

In their study on the prediction of stature from somatometry of the skull Chiba and Terazawa (27) reported that female subjects had shown smaller correlation coefficients than males. Chiba and Terazawa also emphasized that the morphological structure of the head, especially the antero-posterior diameter, changed in relation to the age of the individual. The authors reported a significant increase in the correlation coefficients when individuals over 70 years were excluded from the sample (27).

The studies on body height prediction from cephalometric values are limited, and the results reported in these studies were highly inconsistent. Krishan and Kimura in their study on Koli male adolescents from North India reported a high correlation between cephalo–facial measurements and stature (30). The correlation coefficients were between 0.773 and 0.265. However, the Koli

TABLE 4—*The correlations between the head dimensions and stature for the components of breadth/length index (100 × head breadth/head length).*

Variable	General Population (n = 286)	Breadth/Length Index (100 × Head Breadth/Head Length)			
		Dolicocephalic (n = 22)	Mesocephalic (n = 81)	Brachycephalic (n = 95)	Hyperbrachycephalic (n = 86)
MHL	0.229**	0.139	0.138	0.268**	0.120
MHB	0.012	0.034	0.159	0.241**	0.065
MFD	0.012	0.298	0.087	−0.566	0.039
HC	0.196**	0.280	0.187	0.264**	0.023
MHH	0.162**	0.076	0.157	0.182	0.095

**$p < 0.001$. MHL, maximum head length; MHB, maximum head breadth; MFD, minimum frontal diameter; HC, horizontal circumference; MHH, maximum head height.

TABLE 6—*The correlations between the head dimensions and stature for the components of mean index of height (100 × head height/half the sum of breadth and length).*

Variable	General Population (n = 286)	Mean Index of Height		
		Low (n = 211)	Medium (n = 53)	High (n = 22)
MHL	0.229**	0.177*	0.522**	0.157
MHB	0.012	0.023	−0.032	0.067
MFD	0.012	0.010	−0.147	0.383
HC	0.196**	0.174	0.281*	0.301
MHH	0.162**	0.130	0.419**	0.075

*$p < 0.05$; **$p < 0.001$. MHL, maximum head length; MHB, maximum head breadth; MFD, minimum frontal diameter; HC, horizontal circumference; MHH, maximum head height.

TABLE 8—*The correlations between face dimensions and stature for the components of zygomandibular index.*

Variable	Zygomandibular Index			
	General Population (n = 286)	Narrow Jaw (n = 191)	Medium Jaw (n = 32)	Wide Jaw (n = 63)
BZB	0.168**	0.135	0.001	0.378**
BGD	0.164**	0.124	0.017	0.462**
FL	0.199**	0.243**	0.029	0.183
SFL	0.122*	0.163*	0.039	0.112

*$p < 0.05$; **$p < 0.001$. BZB, bizygomatic breadth; BGD, bigonial diameter; FL, facial length; SFL, superior facial length.

population is an endogamous group showing homogeneity between individuals. Chiba and Terazawa reported a correlation between head and face dimensions and body height in their study on Japanese cadavers with correlation coefficients changing between 0.32 and 0.53. The authors also indicated that the above-mentioned values had increased after older subjects had been excluded from the sample. In this study, the correlation coefficients were lower than had been expected; between 0.229 and 0.122. It should be noted that the sample that Chiba and Terazawa studied was composed of only 77 subjects, while this study was conducted on 286 subjects. In another study on 358 young male subjects by Introna, body height was predicted from head breadth and antero-posterior length (34). The authors reported that the correlation coefficients ranged from 0.38 to 0.60, and the formulae calculated in the study were applicable in the forensic field. Kalia et al. (35) tried to estimate stature using odontometry and skull anthropometry. Correlation coefficients were between 0.56 and 0.38 and were close to those of Chiba and Terazawa, but only when combined data were used. When male and female individuals were evaluated separately, a significant decrease was observed in the correlation coefficients. For male subjects, they changed between 0.20 and 0.13, which is similar to the results of this study. However, in their study on 220 autopsied cases, Sarangi et al. (31) evaluated correlations between maximum transverse length and circumference of the skull and stature and reported that the correlation coefficient of stature for those parameters was insignificant for a reliable estimation of stature.

The most surprising results reported on this subject were those of Krishan and Kimura. Correlation coefficients calculated by the authors were higher than those of all the other researchers who carried out studies on stature estimation from craniofacial dimensions. Correlation coefficients for head circumference and MHL were 0.781 and 0.775, respectively, and those values are close to the coefficients for the long bones. Long bones of limbs give the most reliable estimations for body height. Mall et al. (36) in their study on stature estimation from the long bones of the arm reported the correlation coefficient for the maximum length of the ulna as 0.71 and for the radius as 0.74. Those values are lower than those of the HC of the head and MHL calculated by Krishan and Kimura. When compared with the upper extremity, long bones of lower limbs give more accurate results related with stature estimation. Munoz et al. (37) reported in their study that the femur was the most accurate predictor of stature, and the correlation coefficient value was 0.851. Although this value is higher than that calculated by Krishan and Kimura for head circumference, the correlation coefficients are quite similar to each other.

The results of this study indicate that head length gives the best correlation coefficients with body height. Similar to this study, Patil and Mody (28) also reported head length as the best predictor for stature. Chiba and Terazawa (27) achieved the best estimations with regard to the sum of antero-posterior head length and head circumference. Krishan and Kimura who had reported high correlation coefficients between cephalometric measurements and stature indicated that head circumference was the best indicator for predicting stature (16,30). On the other hand, Ryan and Bidmos (29) indicated that basi-bregmatic height for male subjects and BZB for females gave the highest correlation coefficients for stature. Similarly, in this study, a significant correlation between stature and BZB was found, but its reliability was very low (0.168).

As the correlation coefficients were notably low, a multiple regression formula, reported below, was calculated for a more reliable stature estimation. However, this formula is not reliable enough too

$$\begin{aligned}\text{Stature} = {} & 1163.853 + 1.586\,\text{maximum head length} \\ & + 1.65\,\text{face length}\,t \\ & + 0.805\,\text{bigonial diameter}\end{aligned}$$

$$r = 0.322\,(R^2 = 0.104)$$

The morphological structure of head and face shows variations worldwide according to ethnic background, environmental factors, and even to the habitudes of the populations. Structural differences

TABLE 7—*The correlations between the face dimensions and stature for the components of facial index (100 × face height/BZB).*

Variable	General Population (n = 286)	Facial Index				
		Hypereuryprosope (n = 34)	Euryprosope (Broad or Low) (n = 82)	Mesoprosope (Medium) (n = 91)	Leptoprosope (Narrow or High) (n = 50)	Hyperleptoprosope (n = 29)
BZB	0.168**	0.176	0.220*	0.127	0.298*	0.222
BGD	0.164**	0.239	0.085	0.159	0.379**	0.042
FL	0.199**	0.296	0.279*	0.104	0.394**	0.260
SFL	0.122*	0.033	0.126	0.023	0.312*	0.173

*$p < 0.05$; **$p < 0.001$. BZB, bizygomatic breadth; BGD, bigonial diameter; FL, facial length; SFL, superior facial length.

could also be observed between the individuals in the same population. As correlation coefficients between head and face dimensions and body height can change according to different head and face types, we set out to evaluate the correlation coefficients for individuals who belong to a certain head or face type. However, when correlation coefficients were evaluated after the sample was divided into subgroups according to head or face types, no significant changes were observed when compared with the entire sample.

Especially in endogamous populations, correlation coefficients for head and face dimensions and stature are high, and thus applicable for identification in the forensic field. The Gujjars of North India, on whom Krishan studied the reliability of the stature estimation based on craniofacial dimensions, is an endogamous population marrying within the same caste. This is most likely the main reason for the high correlation coefficients calculated by the researchers. Studies on Japanese, South African, and Italian populations indicated significant correlations between cephalometric measurements and stature. The Turkish population is a considerably mixed population composed of individuals from several ethnic origins. This is probably the main reason for the low correlation coefficients between head and face dimensions and stature.

In conclusion, cephalometric measurements are not reliable variables for stature estimation in the Turkish population.

References

1. Lundy JK. The mathematical versus anatomical methods of stature estimate from long bones. Am J Forensic Med Pathol 1985;6(1):73–6.
2. Raxter MH, Auerbach BM, Ruff CB. Revision of the Fully technique for estimating statures. Am J Phys Anthropol 2006;130:374–84.
3. Steward TD. Essentials of forensic anthropology, 1st edn. Springfield, IL: C. C. Thomas, 1979; 145–76.
4. Ousley S. Should we estimate biological or forensic stature? J Forensic Sci 1995;40:768–73.
5. Bhatnagar DP, Thapar SP, Batish MK. Identification of personal height from the somatometry of the hand in Punjabi Males. Forensic Sci Int 1984;24:137–41.
6. Manouvrier L. Determination de la taille d'apres les grands os de membres. Rev Mensuelle de l'Ecole d'Anthropologie 1892;2:227–33.
7. Pearson K. Mathematical contributions to the theory of evolution: on reconstruction of the stature of prehistoric races. Philos Trans R Soc Lond 1899;192:169–224.
8. Krogman WM. The human skeletons in forensic medicine, 2nd edn. Springfield, IL: Charles C. Thomas Publ., 1978;302–52.
9. Formicola V, Franceschi M. Regression equations for estimating stature from long bones of early Holocene European samples. Am J Phys Anthropol 1996;100:83–8.
10. Prothro JW, Rosenbloom CA. Physical measurements in an elderly black population: knee height as the dominant indicator of stature. J Gerontol 1993;48(1):15–8.
11. Mohanty NK. Prediction of height from percutaneous tibial length among Oriya population. Forensic Sci Int 1998;98:137–41.
12. Duyar I, Pelin C. Body height estimation based on tibia length in different stature groups. Am J Phys Anthropol 2003;122:23–7.
13. Chumlea WC, Rochhe AF, Steinbaugh ML. Estimating stature from knee height for persons 60 to 90 years of age. J Am Geriatr Soc 1985;30(2):116–20.
14. Saxena SK. A study of correlations and estimating of stature from hand length, hand breadth and sole length. Anthropol Anz 1984;42:271–6.
15. Wilbur AK. The utility of hand and foot bones for the determination of stature in a prehistoric population from west-central Illinois. Int J Osteoarchaeol 1998;8:180–91.
16. Kimura K. Estimation of stature in children from second metacarpal measurements. Z Morphol Anthropol 1992;79(1):11–20.
17. Sağır M. Estimation stature from X-rays of metacarpals in the Turkish population. Anthropol Anz 2006;64(4):377–88.
18. Meadows L, Jantz RL. Estimation of stature from metacarpal lengths. J Forensic Sci 1992;37(1):147–54.
19. Himes JH, Yarbrough C, Martorell R. Estimation of stature in children from radiographically determined metacarpal length. J Forensic Sci 1977;22(2):452–5.
20. Bidmos M. Adult stature reconstruction from the calcaneus of South Africans of European descent. J Clin Forensic Med 2006;13(5):247–52.
21. Bidmos M, Asala S. Calcaneal measurement in estimation of stature of South African blacks. Am J Phys Anthropol 2005;126(3):335–42.
22. Holland TD. Brief communication: estimation of adult stature from the calcaneus and talus. Am J Phys Anthropol 1995;96(3):315–20.
23. Zverev Y, Chisi J. Estimating height from arm span measurement in Malawian children. Coll Antropol 2005;29(2):469–73.
24. Zverev Y. Relationship between arm span and stature in Malawian adults. Ann Hum Biol 2003;30(6):739–43.
25. Pelin C, Duyar I, Kayahan EM, Zağyapan R, Ağıldere EM. Body height estimation based on dimensions of sacral and coccygeal vertebrae. J Forensic Sci 2005;50:294–7.
26. Giroux CL, Wescott DJ. Stature estimation based on dimensions of the bony pelvis and proximal femur. J Forensic Sci 2008;53(1):65–8.
27. Chiba M, Terazawa K. Estimation of stature from somatometry of skull. Forensic Sci Int 1998;97:87–92.
28. Patil KR, Mody RN. Determination of sex by discriminant function analysis and stature by regression analysis: a lateral cephalometric study. Forensic Sci Int 2005;147:175–80.
29. Ryan I, Bidmos MA. Skeletal height reconstruction from measurements of the skull in indigenous South Africans. Forensic Sci Int 2007;167:16–21.
30. Krishan K. Estimation of stature from cephalo-facial anthropometry in North Indian population. Forensic Sci Int 2008;3:52e1–e6.
31. Sarangi AK, Dadhi B, Mishra KK. Estimating of stature from adult skull bone. J Ind Acad Forensic Med 1981;182:24–6.
32. Olivier G. Practical anthropology. Springfield, IL: Charles C. Thomas, 1969.
33. Cameron N, Hiernaux J, Jarman S, Marshall WA, Tanner JM, Whitehouse RH. Anthropometry. In: Weiner JS, Lourie JA, editors. Practical human biology. London: Academic Press, 1981;25–52.
34. Introna F Jr, Di Vella G, Petrachi S. Determination of height in life using multiple regression of skull parameters. Boll Soc Ital Biol Sper 1993;69(3):153–60.
35. Kalia S, Shetty SK, Patil K, Mahima VG. Stature estimation using odontometry and skull anthropometry. Indian J Dent Res 2008;19(2):150–4.
36. Mall G, Hubing M, Büttner A, Kuznik J, Penning R, Graw M. Sex determination and estimation of stature from long bones of the arm. Forensic Sci Int 2001;117:23–30.
37. Munoz JL, Linares-Iglesias M, Suarez-Penaranda J, Mayo M, Miguens X, Rodrigez-Calvo MS, et al. Stature estimation from radiographically determined long bone length in a Spanish population sample. J Forensic Sci 2001;46(2):363–6.

Additional information and reprint requests:
Ayla Kürkçüoğlu, M.D., Ph.D.
Department of Anatomy
Faculty of Medicine
Baskent University
06530 Baglıca
Ankara
Turkey
E-mail: kayla@baskent.edu.tr

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01460.x
Available online at: interscience.wiley.com

TECHNICAL NOTE
CRIMINALISTICS

Catherine M. Grgicak,[1] *Ph.D.; Zena M. Urban,*[1] *B.S.; and Robin W. Cotton,*[1] *Ph.D.*

Investigation of Reproducibility and Error Associated with qPCR Methods using Quantifiler® Duo DNA Quantification Kit*

ABSTRACT: Reproducibility of quantitative PCR results is dependent on the generation of consistent calibration curves via accurate volume transfers and instrument performance. A review of 14 standard curves, using two different QuantDuo® standard DNA lots, showed variability of cycle threshold values between assays were larger than those of the Internal PCR Control (IPC). This prompted a set of experiments designed to determine the source of variability. Results showed that error introduced during DNA addition to the plate resulted in little variation. A comparison of seven independent series demonstrated cycle threshold variation between dilutions was larger than the variation expected from repeated samples. Modeling the influence of pipette errors on dilution series accuracy indicated that a more rigorous approach to external calibration curve production is required and showed that improvement in calibration curve stability is expected if the pipette conditions are carefully chosen and/or a single validated curve is utilized as the calibrator.

KEYWORDS: forensic science, DNA quantification, quantitative polymerase chain reaction, real-time polymerase chain reaction, propagation of error, reproducibility

The number of tools available for processing DNA samples continues to increase, providing opportunities to produce more definitive results from difficult samples. One of the most important methodologies available is that of quantitative PCR (qPCR), also known as real-time PCR. With the advent of qPCR multiplex systems, which allow measurement of both total human and total male DNA (1,2), the information gathered during quantification can be used to direct the first steps in forensic DNA analysis.

Human-specific qPCR is one of the most sensitive and versatile techniques for DNA quantification and is commonly used in forensic laboratories (3–5). It is able to provide information on DNA concentration—both total and male—while use of an Internal PCR Control (IPC) provides information on the presence of inhibitors (6). However, to make maximum use of the data, a complete understanding of the ways in which qPCR variability impacts results is essential.

Inconsistent qPCR results can originate from a variety of sources, including but not limited to, variable calibrant concentrations (7), inherent PCR inefficiencies (8), and pipette errors. The variability between and within quantification methods has an immense impact on interlaboratory results and is large enough to warrant development of a NIST Standard Reference Material (SRM 2372) (9). The SRM is designed to give forensic DNA laboratories a method to evaluate each new quantification calibrant lot and give forensic analysts the ability to confirm standard DNA concentrations (10). Although SRM 2372 aids in proper quantification of calibrants, it was not intended for use as a routine calibrant or quality control material.

[1]Biomedical Forensic Sciences, Boston University School of Medicine, Boston, MA.

*Portions of the results were presented at the 19th International Symposium for Human Identification, October 13–16, 2008, in Hollywood, CA.

Received 3 Mar. 2009; and in revised form 21 Aug. 2009; accepted 6 Sep. 2009.

For any given assay, increased sensitivity introduces the potential for a decrease in reproducibility. This is particularly true when quantification is performed by utilizing an external calibration curve generated by a dilution series (11–14). Such a dilution cannot be prepared without introduction of error, which is then propagated from high to low concentrations. For qPCR applications, the inability to precisely and accurately pipette small volumes is exacerbated by the logarithmic relationship between cycle threshold and initial concentration of each standard sample (8,15).

This article represents the first detailed study designed to evaluate error in forensic qPCR analysis by assessing whether utilization of an external calibration curve is robust and results in curves that do not significantly differ between assays. The primary objective of this work is to elucidate the origin and effect deviations of slope and *y*-intercept have on final results by employing ABI's Quantifiler® Duo kit (Applied Biosystems Inc., Foster City, CA). Reproducibility of standard curves, pipette volumes, and the influence these have on derived DNA concentrations were assessed and compared to models predicting expected deviation. The suitability of utilizing a single validated curve as external calibrator rather than the generation of new curves for each assay is considered.

Materials and Methods

All aspects of the study were conducted in compliance with ethical standards set out by the Institutional Review Board of Boston University School of Medicine—Protocol H-26187.

Standard Reproducibility Study

Quantifiler® Duo standard curves generated over a period of 7 months, using two different Quantifiler® Duo standard DNA lots (Lot # 0711001—Lot A and 0803002—Lot B) for both human

(RPPH1) and Y (SRY) targets were compiled to examine reproducibility of the calibration curves. Each curve, ranging from 50 to 0.023 ng/μL, was generated in duplicate in a 25-μL reaction volume following the manufacturers recommended protocol (16). The series was created by first diluting the stock standard (200 ng/μL) to 50 ng/μL by adding 10 μL of stock to 30 μL of dilution buffer provided in the kit. Seven threefold serial dilutions followed, where 10 μL of each successive sample was diluted in 20 μL of buffer. qPCR was performed using an Applied Biosystems Prism® 7500 Sequence Detection System (Applied Biosystems Inc) in 96-well microtiter plates under 9600 emulation conditions.

To measure the inherent variations originating from reaction mix and/or instrument combination, two series of calibration data were compared. Series 1 consisted of a master mix/DNA mixture, where the combination was made up in a volume large enough for four identical reactions. That is, 8 μL of each DNA standard was added to 92 μL of master mix and vortexed. Twenty-five microliters of this mixture was subsequently added to four wells of the optical 96-well microtiter plate. A second series (Series 2) of calibration standards was tested in quadruplicate by utilizing the manufacturers recommended protocol, where the DNA was pipetted individually into each well, following master mix addition. Comparison of the variation in the cycle threshold values for each series provided information on inherent minimum deviation of the assay versus disparities originating from pipette error.

To measure disparities originating from dilution series production, variability of cycle thresholds between a set of four replicates from a single serial dilution was compared to that of six independent serial dilutions. The manufacturers recommended protocol was used to create the seven separate series, and the four replicates were randomly distributed within the 96-well microtitre plate (i.e. in columns 1, 3, 6 and 9).

Visible spectroscopy (Genesys 10S; ThermoScientific Inc, Waltham, MA) was used to determine the concentration of each standard DNA lot where it was assumed that one absorbance unit corresponded to 50 ng/μL of double-stranded DNA. A_{260}:A_{280} was assessed to ensure there was no protein contamination. As the Genesys 10S is a single-beam instrument, a blank sample containing the dilution buffer was run prior to the sample and the background was subtracted. The scan ranged from 400 to 200 nm. Triplicates of Lot B dilutions were run to determine the error associated with UV-Vis results.

Modeling the expected pipette variation during dilution series production was explored by applying the theory of propagation of random and systematic error. In this scenario, 10 to 100-μL pipette volume bias of ±0.8 μL and uncertainty of ±0.3 μL were propagated throughout a series and expected deviations charted. Other scenarios which typically may be encountered in a forensic laboratory were also examined, and the resultant calibration curves plotted. These scenarios included pipette volume bias of ±0.5 μL, ±1 μL and ±2 μL for either the buffer or the DNA transfers.

Evaluation of qPCR Methodologies

High molecular weight, single source male and female samples were extracted from whole blood using phenol/chloroform extraction and alcohol precipitation. All reagents were purchased from Sigma-Aldrich (Sigma-Aldrich Corporation, St. Louis, MO) unless otherwise noted. For each 750 μL whole blood sample, 750 μL of 1× SSC (0.15 M NaCl, 0.015 M tri-sodium citrate) was added to the sample tube, bringing the total volume to 1.5 mL. After gently mixing and centrifuging the sample, 1.0 mL of supernatant was removed, and the remaining pellet was resuspended in 1.0 mL of 1× SSC. Following a second centrifugation, 1.4 mL of supernatant was removed and discarded. To resuspend the pellet and lyse the cells, 375 μL of 0.2 M sodium acetate (pH 7.0), 25 μL of 10% SDS, and 3.2 μL of proteinase K (31.5 mg/mL) were added to the sample and incubated overnight at 56°C. An equal volume of phenol/chloroform was added to the samples, gently mixed, centrifuged, and the organic phase discarded. This process was repeated with the addition of chloroform. DNA precipitation was accomplished with the addition of 2 M sodium acetate, 0.8 μL of 20 mg/mL glycogen, and an equal volume of isopropanol, followed by an overnight incubation at −20°C. The samples were centrifuged for 30 min at maximum angular velocity, and supernatant removed and discarded before the addition of 1 mL of 80% ethanol. The tubes were once again centrifuged, and the supernatant removed and discarded. After allowing the sample to air-dry, the pellet was dissolved in 50 μL of TE buffer (10 mM Tris–HCl, 0.1 mM EDTA, pH 8.0) at 56°C.

The single source and mixture samples were quantified using the Quantifiler® Duo quantification system with the manufacturers recommended protocol (16). DNA mixtures originating from these samples were developed to assess the ability of the assay to quantify male DNA successfully and reproducibly in the presence of female DNA. DNA mixtures of 1:1, 1:2, 1:4, 1:9, 1:19, 1:49, and 1:99 (male:female) were prepared by creating a dilution of female while maintaining the concentration of male DNA at 1 ng/μL. As the dynamic range of the qPCR curve was 50–0.023 ng/μL, the 1:49, and 1:99 mixtures were diluted prior to quantification.

Subsequently, three analysis methods were explored to establish which one decreased variation between assays. The first was the manufacturer's recommended protocol, where a calibration curve is generated per run. The second included the utilization of a corrective factor whereby a standard curve is compared to a reproducible validated curve, and the third utilized a single validated curve. The single source and mixture samples were reanalyzed utilizing each of the methods and results were compared.

Results and Discussion

The most common method used to evaluate an unknown quantity of DNA in qPCR is an external calibration curve derived from a serial dilution. A linear relationship between the cycle threshold (n) and the logarithm of the original concentration of DNA ($C_{j,0}$) is assumed, and a least squares regression to obtain the line of "best" fit ensues.

In the case of ABI's QuantDuo®, a series of standard DNA solutions are obtained by diluting a concentrated DNA standard of 200 to 50 ng/μL. This is followed by seven threefold sequential dilutions down to 0.023 ng/μL for a total of eight calibration standards according to;

$$C_{j+1,0} = \frac{V_j C_{j,0}}{(V_j + V_b)} \quad (1)$$

where V_j is the volume of concentrated DNA, V_b the volume of buffer added to the dilution, j the sample number in the eight sample calibration series. The standards are subjected to 40 temperature cycles and the release of fluorophore from the TaqMan® probes are used to detect the amount of amplified product.

If the amplification efficiency of the target amplicon is 100%, the following relationship is obtained;

$$C_{j,n} = C_{j,0} 2^n \quad (2)$$

where $C_{j,n}$ is the concentration of DNA at cycle n, $C_{j,0}$ is the original concentration of DNA prior to temperature cycling, and j is the sample number. By taking the logarithm of both sides and defining n as the cycle threshold, the familiar linear relationship between n and $\log C_{j,0}$ is obtained

$$\frac{\log C_{j,n}}{\log 2} - \frac{\log C_{j,0}}{\log 2} = n \quad (3)$$

where the y-intercept (b) is $\frac{\log C_{j,n}}{\log 2}$ and the slope (m) is $-\frac{1}{\log 2}$ or -3.32.

According to Eq. 3, if optimal conditions are met and the standards contain accurate DNA concentrations, the y-intercept and slope derived from all calibration curves should exhibit insignificant differences between runs.

Despite this, significant variation has been observed in multiple qPCR-based assays, particularly when the precision study included many runs performed on different days. The H-Quant Human Genomic DNA Quantitation system exhibited slopes of -3.72 ± 0.29 (17), while the developmental validation of Quantifiler® Human showed the coefficient of variation of six samples on three separate runs ranged from 2% to 30% (18). Additionally, Koukoulas et al. showed that standard curves generated using four lots of Quantifiler® Human DNA standard resulted in significant error, whereby concentration differences varied up to a factor of 2. Based on UV-Vis comparisons of various DNA standard lots, the large error in calibration curves was attributed to inconsistencies in standard DNA concentrations between kit lots. However, for two of the three lots, concentration differences were larger than expected given the UV-Vis results, suggesting the presence of additional error (7).

Large variations in calibration curve parameters have also been observed in this laboratory for the Quantifiler® Duo kit. Figure 1 shows 14 Quantifiler® Duo calibration curves generated over a period of 7 months, using two different Quantifiler® Duo standard DNA lots (Lot A and Lot B) for the human (RPPH1) target. The graph also includes the average cycle threshold values obtained from the IPC. This is included in the Quantifiler® Duo reaction mix which is added at a specific concentration by the manufacturer. The IPC is expected to result in similar cycle threshold values from sample to sample assuming, the protocol is followed, the thermal cycler is working properly and there are no inhibitors in the sample. The overall average threshold of the IPC for over 500 samples over the 14 plates was 29.35 with a standard deviation of 0.19.

Of the 14 calibration curves, all R^2 values were >0.98 and the root square error, which describes the error predicted for the y-estimate, ranged from 0.05 to 0.74. Most root square errors (i.e. 12 out of 14) were <0.3. The y-intercepts for the SRY locus were significantly larger than the RPPH1 locus (data not shown), albeit with similar slopes, correlation coefficients, and root square errors. This resulted in SRY standard curves which were "shifted" above the RPPH1 curves, as expected (16). Although an acceptable range for slopes of -3.0 to -3.6 has been recommended by the manufacturer (16), no official recommendation for the y-intercepts has been published.

Figure 1*b* illustrates the slopes and y-intercepts of each of the 14 curves along with standard errors of the parameters. It is of interest to note that parameters with the largest errors also produced the curves with the largest slopes and y-intercepts. However, the 10 curves which resulted in slopes between -3.0 and -3.6 and small standard errors (i.e. <10%) still exhibited a significant spread in y-intercept ranges—between 28.3 and 29.5. The two black points represent the parameters from the curves of two independent dilutions run on the same plate.

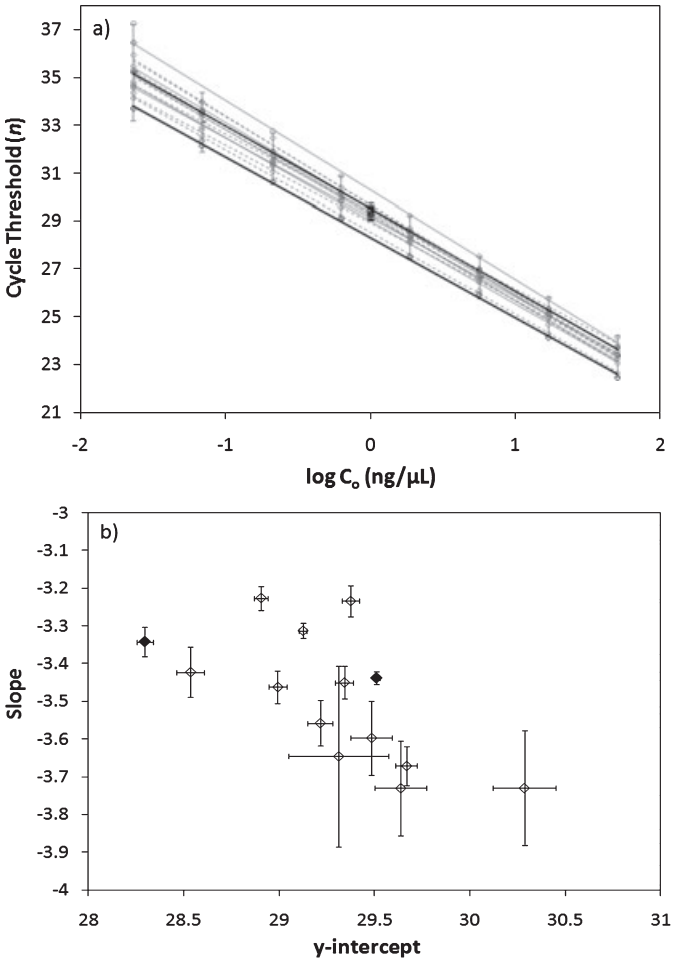


FIG. 1—*(a) A series of 14 standard curves generated over a period of 7 months using two Quantifiler® Duo kit lot numbers. (—) Curves generated from dilution series of standard Lot A. (– –) Curves generated from Standard Lot B. (—) Two curves generated from two separate dilution series from standard lot A and run on the same plate. (□) Average cycle thresholds from the Internal PCR Control of all 14 assays. (b) The slopes and y-intercepts of all curves with the standard errors of each. The solid points are from the two series run on the same plate.*

Figure 1 demonstrates that large calibration curve variations were not dependant on standard lots, but rather on the dilution series. This is further corroborated by the observation that two separate dilution series run on the same plate with similar IPCs resulted in curves that were significantly different (shown as black curves in Fig. 1*a* and black diamonds in Fig. 1*b*). Furthermore, the variation in calibration curves does not coincide with the variation seen in the IPCs which indicates the calibration curve discrepancies are not master mix or instrument dependent. That is, a curve which has a significantly larger y-intercept did not exhibit the same trend in cycle threshold values for the IPCs on the same plate. Additionally, the relative stability of the IPC when compared to the calibrant cycle thresholds between assays indicates the presence of an additional source of error, above and beyond that of expected sample-to-sample variation. The standard deviation of the IPC threshold for all samples was 0.19, while those of the dilution series ranged from 0.41 to 1.00 for the 50–0.023 ng/µL standards, respectively. Additionally, the deviation seen in calibration curves was not solely dependent on standard lot and was corroborated by UV-Vis experiments. UV-Vis showed the concentration of each stock was 126

and 117 ng/μL for Lot A (Lot # 0711001) and Lot B (Lot # 0803002), respectively. The relative error (with a 95% confidence interval) resulting from triplicate dilutions originating from Lot B was 18%. Although the DNA concentrations of the standards were significantly less than the expected 200 ng/μL, they did not significantly differ from one another, demonstrating that in this case, the variation between calibration curves was not lot dependent. However, the sizable discrepancy in measured versus expected stock concentrations illustrates the need for a NIST Standard Reference Material.

The consistency of cycle thresholds of the IPCs between plates, similarity of standard concentrations between lots, high R^2 values, and low root standard errors suggest that imprecision was not the only cause of the large deviations between curves, but a consistent bias propagated throughout the series that was introduced during dilution series preparation.

To explore this further, reproducibility studies designed to evaluate the influence pipette imprecision has on curve generation were undertaken. In this scenario, a series of four calibration standards was prepared by premixing master mix and DNA prior to aliquoting into the 96-well plate (Series 1). These results were compared to the ones obtained using four replicates following the recommended protocol (Series 2).

Figure 2 depicts the RPPH1 and SRY responses obtained for each series. The standard deviation of cycle thresholds for the human target increased with decreasing concentration and ranged from 0.02 for the 50 ng/μL sample to 0.75 for the 0.023 ng/μL sample for Series 1. Series 2 deviations were similar and ranged from 0.01 (50 ng/μL) to 0.69 (0.023 ng/μL). The SRY spread followed the same trend with threshold variations ranging from 0.02 to 0.82 and 0.04 to 0.41 for Series 1 and 2, respectively. As the concentration decreased, the variation increased, suggesting a weighted regression analysis would provide better estimates of the linear parameters. Work previously performed in this laboratory suggests that, although possibly a more prudent regression model, weighted regression does not significantly impact final results. The standard deviations of the cycle threshold values obtained in this laboratory were small for both quadruplicate series and are similar to those obtained by Green et al. (18). The regression lines are shown and are indistinguishable between series, suggesting an insignificant difference between sample responses. An increase in the standard deviation with dilution is observed for both series, but did not differ between them; suggesting random pipette error originating from sample addition to the amplification plate was an insignificant source of error.

As the aforementioned data demonstrate that the irreproducibility of calibration curves does not necessarily originate from differences in stock DNAs or random pipette errors during amplification setup, errors generated during production of the dilution series were investigated. Figure 3*a* shows a series of seven dilutions run on the same plate using the same standard lot (Lot B). The first series was run in quadruplicate (solid lines), and variability of cycle thresholds between a single serial dilution versus an additional six series was evaluated (Fig. 3*b*). The linear parameters were similar to those of Fig. 1, and the IPCs were also within the same range. The variability of cycle thresholds for each of the standards was significantly larger than those generated from the equivalent four replicates, suggesting the cycle thresholds between dilution series varies much more and independently of cycle thresholds of repeated samples.

Variability in qPCR concentrations is drastically influenced by the ability of the laboratory and/or analysts to create a consistent and certifiable dilution series. Generation of such series' will be

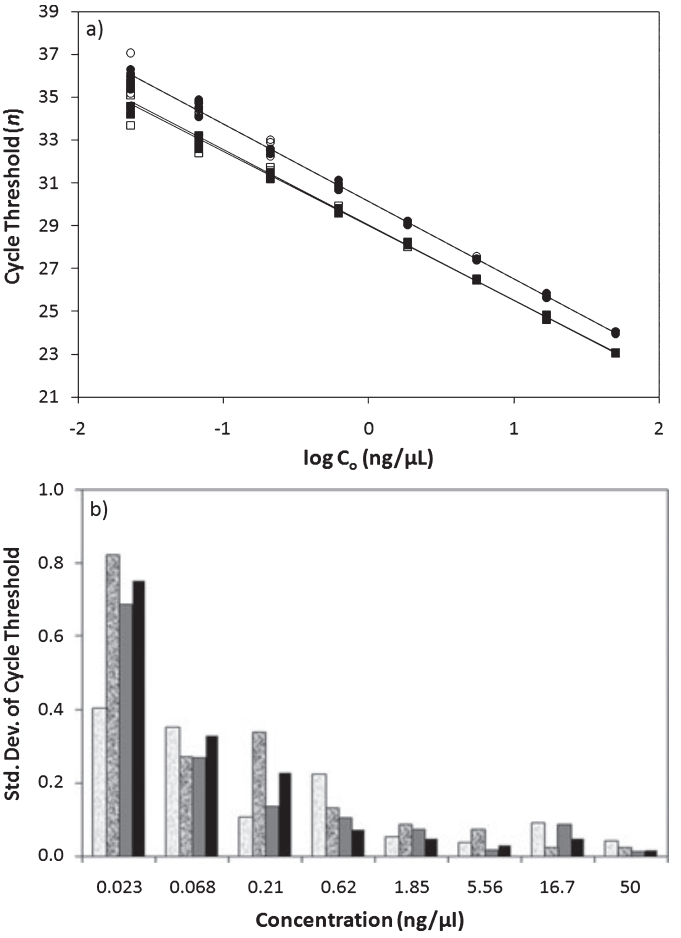


FIG. 2—*Reproducibility of quantitative PCR standard curves. (a) Cycle threshold versus the logarithm of the original concentration for premixed master mix/DNA (Series 1) for (□) human (RPPH1) and (○) Male (SRY) DNA. Cycle threshold versus the logarithm of concentrations for results obtained using the manufacturer's recommended protocol (Series 2) for (■) human (RPPH1) and (•) Male (SRY). (b) The standard deviation (Std. Dev.) for human (RPPH1) (black bar) Series 1 and (gray bar) Series 2. The standard deviation for Male (SRY) DNA for (dark pattern) Series 1 and (light pattern) Series 2.*

dependent on the accuracy of the volume transfer. To establish the influence pipette discrepancies have on standard concentrations, an analysis of pipette error in qPCR during dilution series production was evaluated.

Analysis of Pipette Error in qPCR During Dilution Series Production

The external calibration method implicitly assumes all error occurs in the *y*-direction and these errors are independent of analyte concentration. That is, all error is expected to originate in the cycle threshold—or fluorescent measurement thereof—and not in the DNA concentrations of standards. However, errors in the *x*-direction may result in concentrations which significantly differ from expected values. Furthermore, serial dilutions of standard DNAs from high to low concentrations are expected to propagate error in the *x*-direction. These compounding errors would result in standard curves which are less reproducible between qPCR assays.

As the lowest concentration of Quantifiler® Duo standard is diluted in seven steps from the highest by the same factor (i.e. 3) each time, propagation of both bias and uncertainty occurs. According to the theory of propagation of systematic error (19,20),

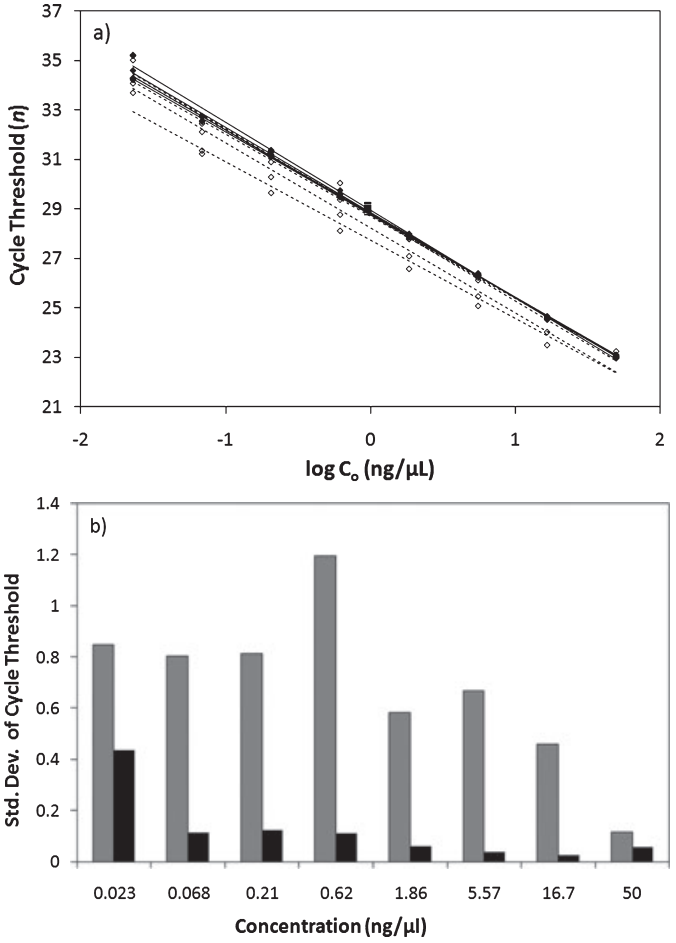


FIG. 3—*Reproducibility of human (RPPH1) quantitative PCR standard curves obtained from seven separate series run on the same plate. (a) Cycle threshold versus the logarithm of the original concentration of the (♦) first series run in quadruplicate and (◇) six independent series. (—) The trendlines of the first series and (– –) the trendlines from the six independent series. (b) The human (RPPH1) standard deviation (Std. Dev.) for (black bar) the first series run in quadruplicate and (gray bar) the six independent series.*

$$\varepsilon_{\mathrm{f}} = \sum_{i=1}^{K}\left(\frac{\partial f}{\partial p_i}\varepsilon_{\mathrm{p_i}}\right) \tag{4}$$

By substituting Eq. 1 into Eq. 4, partially differentiating with respect to each parameter (i.e. V_{j}, $V_{\mathrm{j+1}}$ and $C_{\mathrm{j,0}}$) and assuming errors are independent, the following equation, representing the propagation of volume bias from one sample to another, is obtained.

$$\varepsilon_{C_{\mathrm{j+1,0}}} = \frac{1}{\left(V_{\mathrm{j}}+V_{\mathrm{b}}\right)^2}\left(C_{\mathrm{j,0}}V_{\mathrm{b}}\varepsilon_{V_{\mathrm{j}}} + \left(V_{\mathrm{j}}^2 + V_{\mathrm{j}}V_{\mathrm{b}}\right)\varepsilon_{C_{\mathrm{j,0}}} - C_{\mathrm{j,0}}V_{\mathrm{j}}\varepsilon_{V_{\mathrm{b}}}\right) \tag{5}$$

where $\varepsilon_{C_{\mathrm{j+1}}}$ is the error associated with the concentration of dilution *j*. As the serial dilution is generated, this error is proliferated throughout the DNA standards. Equation 5 shows errors associated with DNA concentration ($C_{\mathrm{j,0}}$) and pipette volumes (V_{j} and V_{b}) play a role in the accuracy of calibration curves. The effect sequential error, originating from a bias of ±0.8 µL associated with a 10–100 µL pipette, has on standard concentrations is shown in Table 1.

As ±0.8 µL is at the limit of the manufacturers recommendations for accuracy during pipette calibration for a 10 to 100-µL pipette (21), the erroneous DNA concentrations in Table 1 represent the maximum allowable bias for the pipette volumes, which affects both the numerator and denominator of the volumes (V_{j} and V_{b}) of Eq. 1. By doubling the volume, the relative error decreased by half that of the original, suggesting use of larger volumes would alleviate issues originating from pipette bias. Other strategies designed to decrease the impact of incorrect volume transfers should also be considered. Such strategies may include those which are intended to decrease pipette volume "drift" over time by using a fixed volume pipette allocated for qPCR transfers, calibrating the pipette at that fixed volume and calibrating more often. A number of these strategies will be examined in detail below.

By arbitrarily choosing a threshold concentration (i.e. 5.38×10^8 ng/µL), one could calculate the cycle threshold (*n*) for each concentration in the standard dilution series via Eq. 3. For example, for a $C_{\mathrm{j,0}}$ of 50 ng/µL the cycle threshold is expected to be 23.36. However, if the actual concentration was 48 ng/µL, the cycle threshold would be 23.42. The thresholds would continue to deviate in a more pronounced manner as the dilution series continued as shown in Fig. 4*a*.

Figure 4*a* also shows the propagated uncertainty associated with pipette discrepancies for the aforementioned scenario and is illustrated as the error bars in the *y*-direction. According to the theory of propagation of random error (19,20),

$$\sigma_{\mathrm{f}}^2 = \sum_{i=1}^{K}\left(\frac{\partial f}{\partial p_{\mathrm{i}}}\sigma_{\mathrm{p_i}}\right)^2 \tag{6}$$

where p_{i} denotes a given parameter, $\sigma_{\mathrm{p_i}}$ is the error in p, σ_{f} the uncertainty produced by $\sigma_{\mathrm{p_i}}$. A similar approach to that taken with bias gives,

$$\sigma_{C_{\mathrm{j+1,0}}} = \frac{1}{\left(V_{\mathrm{j}}+V_{\mathrm{b}}\right)^2}\sqrt{\left(C_{\mathrm{j,0}}V_{\mathrm{b}}\sigma_{V_{\mathrm{j}}}\right)^2 + \left(\left(V_{\mathrm{j}}^2 + V_{\mathrm{j}}V_{\mathrm{b}}\right)\sigma_{C_{\mathrm{j,0}}}\right)^2 + \left(C_{\mathrm{j,0}}V_{\mathrm{j}}\sigma_{V_{\mathrm{b}}}\right)^2} \tag{7}$$

where $\sigma_{C_{\mathrm{j+1}}}$ is the uncertainty associated with the concentration of dilution *j*. Sequential error originating from a uncertainty of ±0.3 µL associated with a 10 to 100-µL pipette (21) increased with decreasing concentration, but was relatively small and indistinguishable from the points.

Although errors originating from a single calibrated pipette have some effect on reproducible and accurate curves, it does not account for the large irreproducibility between calibration curves observed in Fig. 1. If a single calibrated pipette was used, the spread in the *y*-intercepts would not exceed 0.4. Additionally, the slope would only be expected to exhibit a difference of 0.14. However, variations depicted previously in Fig. 1 resulted in a *y*-intercept and slope spread of 2.0 and 0.5, respectively.

Figure 4*b* shows curves which have been generated using the methodology described above. In this case, the bias is ±0.8 µL, but pipette volumes have been doubled as seen in Table 1. Figure 4*c* shows the result of using a more targeted pipette (i.e. 5–50 µL) of which the allowable bias is ±0.5 µL. Figure 4*d* exhibits the resultant curves generated from a dilution series where the error is present only in the most concentrated sample. In this scenario there was a transfer error of ±2 µL during the creation of the first standard (i.e. 50 ng/µL). Here, the slopes remain constant, while only the *y*-intercept changes. Figures 4*e* shows curves which are generated by a pipette which is out of calibration (i.e. bias of ±2 µL)

TABLE 1—*The effect of sequential error on DNA concentration originating from a bias of ±0.8 μL associated with a 10 to 100-μL pipette.*

(*j*)	V_j(μL)	V_b(μL)	Expected Conc. (ng/μL)	Conc. After a Bias of −0.8 μL	Conc. After a Bias of +0.8 μL	Percent Deviation After a Bias of −0.8 μL	Percent Deviation After a Bias of +0.8 μL
0			200	200	200	0	0
1	10	30	50	48	52	4.0	−4.0
2	10	20	16.7	15.6	17.9	6.6	−7.2
3	10	20	5.56	5.04	6.07	9.4	−9.2
4	10	20	1.85	1.63	2.07	11.9	−11.9
5	10	20	0.617	0.527	0.708	14.6	−14.7
6	10	20	0.205	0.170	0.241	17.1	−17.6
7	10	20	0.068	0.055	0.082	19.1	−20.6
8	10	20	0.023	0.018	0.028	21.7	−21.7
1	20	60	50	49	51	2.0	−2.0
2	20	40	16.7	16.1	17.2	3.6	−3.0
3	20	40	5.56	5.30	5.81	4.7	−4.5
4	20	40	1.85	1.74	1.96	5.9	−5.9
5	20	40	0.617	0.572	0.662	7.3	−7.3
6	20	40	0.205	0.188	0.224	8.3	−9.3
7	20	40	0.068	0.062	0.075	8.8	−10.3
8	20	40	0.023	0.020	0.025	13.0	−8.7

and Figure 4*f* illustrates the curves originating from the use of two pipettes; one that has a +0.8 μL bias for the pipette which is used to transfer DNA and −0.8 μL bias for the pipette used to transfer the buffer. The top curve is the opposite scenario, where the DNA pipette is inaccurately transferring by a volume of −0.8 μL.

The results presented in Fig. 4 indicate that a more rigorous approach to external calibration curve production is required and a significant improvement in calibration curve consistency may be expected if the pipette conditions are carefully chosen. Such conditions would include the use of one well calibrated pipette with a well-defined error and the use of larger volumes. Given the instability and precariousness of calibration curve generation, an alternative approach would be the utilization of a single validated curve to obtain DNA concentrations of forensic samples. This approach would ensure consistent linear parameters throughout the course of testing and is evaluated below.

Evaluation of qPCR Methodologies

A detailed understanding of the nature and source of calibration curve generation is necessary to appropriately determine DNA quantity within a forensic laboratory, hence allowing for optimal short tandem repeat (STR) results. Fundamental factors, which influence calibration curves, have been shown to occur as a result of transfer errors, inaccurate DNA concentrations, and PCR variations, where the creation of the dilution series is very susceptible to these errors. These errors which are compounded throughout the process require DNA laboratories to rigorously define the allowable variations in qPCR results.

Validation studies of new methodologies for forensic laboratories are necessary according to the Guidelines for a Quality Assurance Program for DNA Analysis by the Technical Working Group on DNA Analysis Methods. Included in these guidelines is the requirement to document the reproducibility and precision of the procedure of interest. Therefore, during the internal validation process, reproducibility of qPCR calibration curves should be examined and a validated equation generated. This validation process will enable the laboratory to obtain the expected slope and *y*-intercept for their instrument. Once these values have been established, three possible approaches to quantification are possible:

1. The Quantifiler® Duo manufacturer's recommended protocol proposes duplicate calibration curves be generated by linear regression for every qPCR assay and suggests the assay be repeated if the slope of the external calibration curve falls outside the suggested range (i.e. −3.0 to −3.6). No suggestions for *y*-intercept ranges are given. The allowable *y*-intercept error is left to the discretion of the laboratories. By choosing an acceptable deviation in DNA quantity (i.e. 20%) a laboratory may calculate the acceptable deviations in slope and *y*-intercept. However, it should be noted that acceptable deviations for each parameter will vary with concentration and are interdependant. As generation of each new calibration curve relies on the accuracy of DNA concentrations in the dilution series, which are highly variable, this method is prone to inter- and intralaboratory discrepancies, resulting in unnecessary error. If the error is large, re-quantification of samples on the plate and therefore unnecessary sample loss is inevitable. Additionally, on a given 96-well plate, 16 wells will need to be reserved for standards, thereby increasing the cost and decreasing sample throughput.
2. The other possibility is running calibration curves for every assay as described above and utilizing the following equation to determine error in slope and *y*-intercept.

$$\frac{C_{j,0}}{xC_{j,0}} = \frac{10^{n - b/-m}}{10^{n - b + \varepsilon_b/-m + \varepsilon_m}} \quad (8)$$

where $C_{j,0}$ is DNA concentration, n cycle threshold, b and m the slope and *y*-intercept respectively, and ε_b and ε_m the errors of each. The x term is the "corrective factor" and can be applied to obtain the quantity of DNA. This equation can be used in lieu of rerunning the assay, thereby saving time and sample and decreasing cost. This methodology requires a recalculation of all samples by the analyst and still requires 16 wells be reserved for standards. This method—or a variation thereof—is currently used in some laboratories, and it should be noted that Eq. 8 (and not a corrective factor of 2^{ε_b}) should be used if this is to be the laboratories method of choice.
3. The third method is one where the calibration curve generated during validation is utilized as the calibrator, and a positive control and IPC are used to measure PCR reproducibility and efficiency. The IPC and positive control are expected to produce cycle threshold values which are similar between runs and may be charted over time. Any significant deviation in the IPCs or positive controls would suggest the instrument and/or run needs

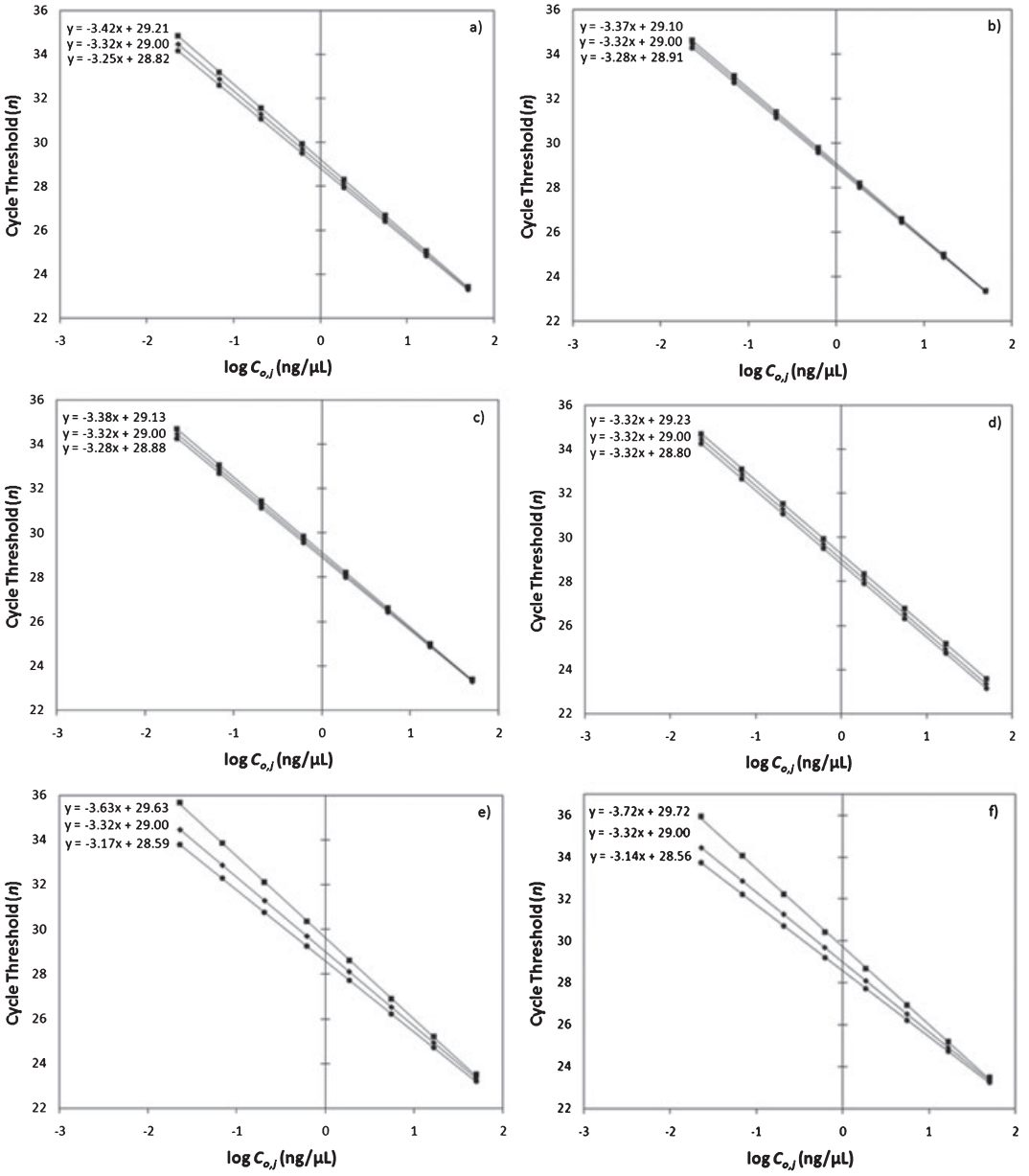


FIG. 4—*(a) Calibration curves obtained when no error (middle curve) and ±0.8 µL bias is present during dilution series production. Resultant cycle threshold deviations from a ±0.3 µL random error is illustrated as error bars in the y-direction. (b) Curves generated when the bias is ±0.8 µL, but pipette volumes have been doubled as seen in Table 1. (c) Bias is ±0.5 µL. (d) Curves generated from a dilution series where a ±2 µL error is present during the creation of the first standard (i.e. 50 ng/µL). (e) Curves generated by a pipette which is out of calibration with a bias of ±2 µL and (f) curves originating from the use of two pipettes; one that has a +0.8 µL bias for the pipette which is used to transfer DNA and −0.8 µL bias for the pipette which is used to transfer the buffer. The top curve is the opposite scenario, where the DNA pipette is inaccurately transferring by a volume of −0.8 µL.*

to be evaluated for failure. This method is one where the calibration curve, generated during validation, is used as the standard curve for subsequent qPCR testing, without making a new set of curves for every plate. Validation of such a curve may include the generation of a number of curves within a plate—preferably using the NIST Standard Reference Material. All points are plotted (with outliers removed) and used to generate the linear parameters. Subsequently, a series of STR amplifications of varying target amounts (i.e. 4–0.0625 ng), using the concentrations generated by the qPCR, will elucidate the appropriate DNA target for STR amplification.

As this method results in the highest throughput of samples as well as reproducible data, it is particularly beneficial for exhaustive DNA testing, or for samples which are inhibited or degraded. Although the positive control and IPC would help analysts determine whether the instrument/reagents are working properly, laboratories may opt to test the curve on a regular basis by running a number of qPCR calibration curves to ensure the parameters (i.e. slope and *y*-intercept) are similar to those previously obtained. It should be noted that the curves will need to be regenerated when there is a QC failure and/or significant instrument modification (i.e. lamp, detector replacement).

Calibration of the curves would ensure the cycle thresholds for each standard, *y*-intercepts, and slopes are consistent over time. Additionally, this method would allow laboratories to create an external calibrator directly from the NIST SRM 2372, eliminating

the need to verify stock DNA concentrations obtained from manufacturers. Therefore, two new sets of dilution series of the standard DNA would not be needed on each plate, thus providing 16 additional wells for sample testing and use of those 16 wells for samples, as opposed to standards. There would be no plate failures because of poorly generated dilutions; hence eliminating unnecessary sample loss because of qPCR "reruns."

It is noted that in all cases, DNA quantities are determined using the calibration curve generated during validation, either through (1) ensuring the curves have similar parameters, (2) by utilizing corrective factors which are derived by comparison to the validated curve, or (3) by using the validated curve itself.

Although the benefits derived by utilization of Method 3 include increased throughput, the increased intralaboratory reproducibility and reduction of unnecessary sample loss are the most desired effects. Figure 5 shows a scenario encountered in this laboratory and one which is commonly encountered, whereby two qPCR assays produced discrepant results.

The solid bars represent resultant DNA concentrations calculated using calibration curves obtained at the time of each run. Each sample was run in quadruplicate and the error bars represent the three standard deviation spread. The slopes and *y*-intercepts were −3.73 and −3.44 and 30.28 and 29.51 for the black and gray, respectively. The large discrepancy between runs and the relatively small error bars, representing amplification variation, supports the previous findings which suggest deviations in the calibration curves—particularly large *y*-intercept deviations—will introduce large DNA concentration discrepancies between qPCR assays. Additionally, modest shifts in *y*-intercept and slope result in DNA concentrations that vary by a factor of 2 and as a result may adversely affect STR amplification. By utilizing Method 3 (i.e. where a single validated curve is used as the calibrator) the same samples produce consistent and reproducible results.

Method 3 was further tested by performing a series of assays which included male: female mixtures of 1: 1, 2, 4, 9, 19, 49, 99. In all cases the male portion was 1 ng/μL while the female concentration varied. Figure 6 shows the concentration fold-differences between the different methods where the black and gray bars represent concentration fold-differences of mixture samples run three times using Method 1 and 3, respectively. Concentration fold-differences is the ratio estimate between runs and was calculated by dividing the DNA concentration by the concentration obtained from the first run.

In general, utilization of Method 3 shows a similar or increased reproducibility between runs, signifying variation in these samples is solely derived from deviations in cycle threshold values of the sample and not the calibration curve. Increases in concentration fold-differences in Fig. 6*b* are larger at higher male:female mixture ratios and do not originate from errors intrinsic in the standard curve. Instead, they result from an increase in cycle threshold variation in the male signal as more female DNA is added (Fig. 7).

Curves generated during assay 1, 2, and 3 resulted in a small *y*-intercept spread and therefore a high degree of reproducibility between these mixture assays. Proper evaluation of DNA concentration of the mixture samples was not hindered by the use of Method 3. In fact, improved results were obtained with this method (Fig. 5) thereby demonstrating Method 3 is not only as effective as the recommended protocol, but an improvement.

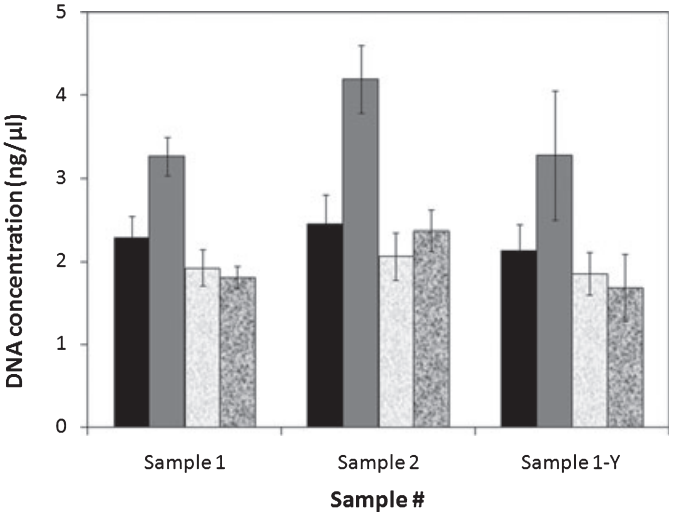


FIG. 5—*DNA concentration of two samples resulting from two different quantitative PCR assays. Total human DNA concentration from assay 1 using curve 1 (black bar), assay 2 using curve 2 (gray bar), assay 1 using validated curve (light pattern) and assay 2 using validated curve (dark pattern). The equations for curve 1, curve 2 and the validated curve were* $n = -3.44 \log C_0 + 29.51$, $n = -3.73 \log C_0 + 30.28$ *and* $n = -3.47 \log C_0 + 29.26$ *respectively for RPPH1. The equations for curve 1, 2 and the validated curve were* $n = -3.42 \log C_0 + 31.16$, $n = -3.18 \log C_0 + 30.31$ *and* $n = -3.37 \log C_0 + 30.16$ *for the SRY locus, respectively.*

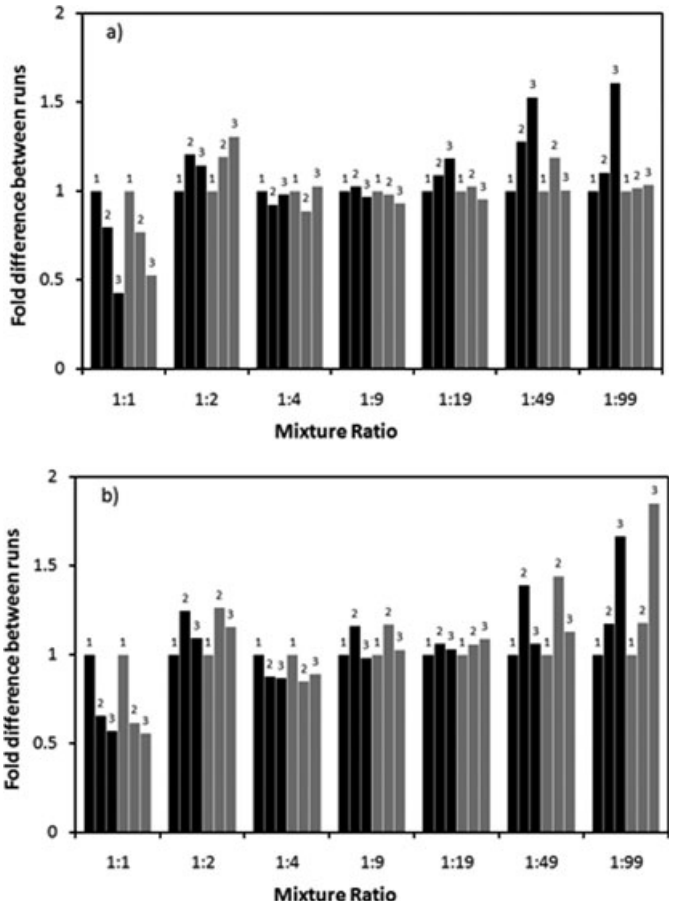


FIG. 6—*Concentration ratio estimates between runs. (a) Calibration curves generated during assay 1, 2 and 3 had a large range in slope (−3.23 to −3.73) and moderate y-intercept range (29.37–29.67) for the human locus. (b) The SRY-slopes and intercepts were −3.57 to −3.79 and 30.61–30.71, respectively. The validated curves used to calculate the gray bars were the same as those in Figure 5.*

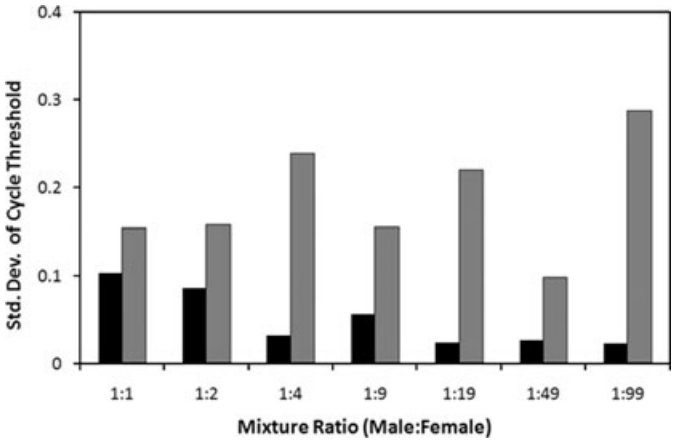


FIG. 7—*Standard deviation of cycle threshold for (black) human (RPPH1) and (gray) Y (SRY) targets for a set of male:female mixtures run in quadruplicate, in which the master mix and DNA were combined and 25 μL subsequently added to the microtiter plate.*

Conclusion

To measure inherent qPCR variations, a single master mix was made with DNA at each concentration tested. This master mix/DNA sample combination was made up in an amount large enough for analysis of four identical reactions. A second set of four reactions was tested by pipetting the sample DNA individually into each well. Comparison of the variation in the cycle threshold (*n*) values for each sample showed variation between repeat samples is small. Variability between dilutions had a larger impact on calibration curve stability. Modeling the propagation of error during dilution series production showed error generated with a single calibrated pipette was not significant and did not account for large calibrant discrepancies observed in this laboratory and the literature. However, the use of two pipettes during dilution series preparation, or the use of an uncalibrated one, was shown to significantly impact curve generation.

Accurate estimates of error associated with DNA concentration require consideration of the logarithmic relationship between cycle threshold and concentration, as well as errors associated with parameters of the standard curve. Three qPCR external calibration methods were discussed, where Method 3, which uses a validated curve as the external calibrator, is recommended because of its ability to increase sample throughput, reproducibility, and eliminate the need to quality check DNA stocks from manufacturers.

Acknowledgments

We thank the anonymous reviewers for their thoughtful and insightful recommendations for improving this manuscript. Funded by the Boston University School of Medicine Biomedical Forensic Sciences Program.

References

1. Horsman KM, Hickey JA, Cotton RW, Landers JP, Maddox LO. Development of a human-specific real-time PCR assay for the simultaneous quantitation of total genomic and male DNA. J Forensic Sci 2006;4:758–65.
2. Nicklas JA, Buel E. Simultaneous determination of total human and male DNA using a duplex real-time PCR assay. J Forensic Sci 2006;51:1005–15.
3. Westring CG, Kristinsson R, Gilbert D, Danielson PB. Validation of reduced-scale reactions for the Quantifiler™ human DNA kit. J Forensic Sci 2007;52:1035–43.
4. Swango KL, Hudlow WR, Timken MD, Buoncristiani MR. Developmental validation of multiplex qPCR assay for assessing the quantity and quality of nuclear DNA in forensic samples. Forensic Sci Int 2007;170:35–45.
5. Nicklas JA, Buel E. Development of an alu-based, real-time PCR method for quantitation of human DNA in forensic samples. J Forensic Sci 2003;48:1–9.
6. Kontanis EJ, Reed FA. Evaluation of real-time PCR amplification efficiencies to detect PCR inhibitors. J Forensic Sci 2006;51:795–804.
7. Koukoulas I, O'Toole F, Stringer P, Oorschot R. Quantifiler™ observations of relevance to forensic casework. J Forensic Sci 2008; 53:135–41.
8. Nordgard O, Kvaloy JT, Farmen RK, Heikkila R. Error propagation in relative real-time reverse transcription polymerase chain reaction quantification models: the balance between accuracy and precision. Anal Biochem 2006;356:182–93.
9. Kline MC, Duewer DL, Redman JW, Butler JM. Results from the NIST 2004 DNA quantitation study. J Forensic Sci 2005;50:571–8.
10. National Institute of Standards and Technology. Certificate of analysis: standard reference material® 2372. Human DNA quantitation standard. Gaithersburg, MD: National Institute of Standards and Technology, 2007.
11. Higgins KM, Davidian M, Chew G, Burge H. The effect of serial dilution error on calibration inference in immunoassay. Biometrics 1998;54:19–32.
12. Racine-Poon A, Weihs C, Smith AFM. Estimation of relative potency with sequential dilution errors in radioimmunoassay. Biometrics 1991;47:1235–46.
13. Lee M-LT. Statistical inference for serial dilution assay data. Biometrics 1999;55:1215–20.
14. Baeza-Baeza JJ, Ramis-Ramos G. Analysis of the sensitivity to the systematic error in least-squares regression models. Anal Chim Acta 2004;515:15–21.
15. Beauchamp JJ, Olson JS. Corrections for bias in regression estimates after logarithmic transformation. Ecology 1973;54:1403–7.
16. Applied Biosystems. Quantifiler® Duo DNA quantification kit. Foster City, CA: Applied Biosystems, 2008.
17. Shewale JG, Schneida E, Wilson J, Walker J, Batzer M, Sinha SK. Human genomic DNA quantitation system, H-quant: development and validation for use in forensic casework. J Forensic Sci 2007;52: 364–70.
18. Green RL, Roisestad IC, Boland C, Hennessy LK. Developmental validation of the quantifiler real-time PCR kits for quantification of human nuclear DNA samples. J Forensic Sci 2005;50:1–17.
19. Miller J, Miller J. Statistics for analytical chemistry, 3rd rev. edn. Chichester, UK: Ellis Horwood Limited, 1993.
20. NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/, March 2008.
21. ThermoElectron Corporation. Finnpipette® focus single channel variable & fixed volume instructions for use. Vantaa: ThermoElectron Corporation, 2008.

Additional information and reprint requests:
Catherine Grgicak, Ph.D.
Boston University School of Medicine
Biomedical Forensic Sciences
72 East Concord Street, Room R806B
Boston, MA 02118
E-mail: cgrgicak@bu.edu

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01427.x
Available online at: interscience.wiley.com

TECHNICAL NOTE
CRIMINALISTICS

Caitlyn Middlestead,[1] *M.S. and John Thornton,*[1] *D.Crim.*

Sensitivity of the Luminol Test with Blue Denim

ABSTRACT: An article appearing in this journal in 2000 suggested that the sensitivity of the luminol test performed on denim fabric is usually no greater than at a 1:100 dilution of blood. This study shows that the luminol test may be unambiguously interpreted at substantially greater dilutions of blood. In this study, four different types of denim were tested by spraying a swatch of fabric with a typical formulation of the luminol reagent. Testing was conducted of dilutions of blood up to 1:1000, all of which showed distinct chemiluminescence. Diluted blood was applied to denim material in the form of a random number. A successful test was obtained only when a "blind" observer, i.e., an observer who was uninformed of the number, correctly reported the number.

KEYWORDS: forensic science, luminol sensitivity, denim substrate, blood, chemiluminescence, presumptive test

Luminol (3-Aminophthalhydrazide or 5-Amino-2,3-dihydrophthalazine-1,4-dione) was first synthesized by Schmitz in 1902 (1,2) and was found to possess chemiluminescent properties by Albrecht in 1928 (3). A more economical synthesis of 3-aminophthalhydrazide was discovered in 1934 by Huntress (4), and it is Huntress who gave the name "Luminol" to the compound. In 1936, Gleu and Pfannstiel (5) discovered that crystalline hemin produced an especially intense reaction when treated with luminol, and almost immediately thereafter, in 1937, the compound was proposed by Specht (6) for use as presumptive blood test.

From the outset, the luminol reaction was viewed as a particularly sensitive, although not specific, test for blood. The typical manner of expressing sensitivity of the reaction is in terms of maximal dilutions of blood, which will give a positive reaction, with the reaction being monitored visually. Proscher and Moody (1) claimed that the test was sensitive to $1{:}10^9$ dilutions of hematin, an extreme claim that has few if any current adherents. Grodsky (7) found the sensitivity to be one part of blood in 1.5×10^6 parts of water. Castellanó (8) gives the sensitivity as 1:300,000 with bloodstains deposited on unspecified cotton cloth. Somewhat higher sensitivities have been reported with instrumental means of monitoring the reaction (9,10). A recent review by Tobe et al. (11) has given the sensitivity to be in the range of 1:100,000.

The maximum sensitivity of the test may be expected to depend on a number of factors—reagent composition, the manner in which the reaction is subjectively followed, and possibly (12) the age of the bloodstain. But whatever the maximum sensitivity is, in no one's experience has it been as low as 1:100.

But then in 2000, the FBI Laboratory reported the sensitivity of the luminol reaction on absorbent materials to be "usually" no greater than 1:100 (13). As this assertion forms the core of the issue addressed in this study, the assertion is given here in its original form:

[With respect to luminol] "stains on absorbent surfaces were detectable usually at no more than a 1:100 dilution. Sensitivity levels were similar to those reported by Frégeau et al."

The thrust of this article by Budowle et al. was in connection with DNA typing, and the use of the terms "usually" and "similar" permit of some negotiation of meaning. But 1:100 is quite different than Castellanó's 1:300,000 or Tobe's 1:100,000. And it suffers from being wrong. In the Budowle et al. article, denim is specifically named as a typical absorbent substrate. The literal 1:100 value appears to be how some forensic workers have now interpreted the sensitivity of luminol toward denim, and the 1:200 value attributed (in error) to Frégeau was picked up and offered in the Tobe et al. review as representing a lower estimate.

The claim that the work of Chantal Frégeau et al. (14) supports the low sensitivity of the luminol reaction as claimed by the FBI Laboratory is misleading. The work of Frégeau et al., as was the work of Budowle et al., was not directed specifically to the sensitivity of luminol, but dealt with the prospects of successful typing by DNA of bloodstains that had been diluted. The Frégeau work is actually quite clear as to the work that was done with luminol. That work was carried out to maximum dilutions of 1:200, but no further. At 1:200 dilutions of blood, Frégeau observed successful luminol reactions. Her work cannot be reasonably interpreted that dilutions in excess of 1:200 would result in negative luminol reactivity; it was simply that she cut off her study at that dilution. Frégeau included blue denim as a test substrate.

The inference of the assertion by the FBI Laboratory that luminol would fail to detect blood on denim if the blood was diluted in excess of 1:100 has profound implications with respect to interpretation. If a luminol test proved negative, would a worker be justified in offering an opinion that blood could in fact be present, but simply at a dilution of greater than 1:100? This study was directed at testing the claim by the FBI Laboratory of the low sensitivity of the luminol test on denim material. It was not the purpose of this study to determine and settle once and for all the maximum sensitivity of the luminol test under each and every set of conditions and with all substrates, but rather to focus on the specific issue of

[1]Forensic Science Graduate Group, University of California, Davis, CA.

Received 24 April 2009; and in revised form 18 July 2009; accepted 31 July 2009.

the FBI Laboratory claim. The extreme limit of sensitivity was not addressed. It is to be expected that the maximum sensitivity will be influenced by reagent composition, conditions of application, possibly by the age of the bloodstain, and even with dark adaptation on the part of the viewer. At extreme dilutions, slight variations could be expected that might alter experimental results. Stated differently, the purpose of the present research was not to establish a definitive number for the maximum sensitivity, but focused rather on the credibility of the suggestion by Budowle et al. that the sensitivity of luminol is lower than has been previously described, with particular reference to an absorbent substrate such as denim.

Materials and Methods

Denim material was chosen for this experiment because both the Frégeau and Bodowle studies used denim as a substrate, typical of a porous material. The present study considered both blue and black denim. Marie (15) has suggested that black denim is a more difficult substrate on which to develop bloodstains than other materials.

Thirteen dilutions of whole blood with deionized water were created and placed onto denim substrates. Whole blood from a 53-year-old Caucasian woman was drawn by venipuncture into EDTA tubes and refrigerated until use. Dilutions were prepared as follows: neat (1:0), 1:100, 1:150, 1:200, 1:250, 1:300, 1:400, 1:500, 1:600, 1:700, 1:800, 1:900, and 1:1000. Blanks consisting of deionized water were also made.

As the luminol reacts with heme in erythrocytes, a hematocrit was performed on the blood to determine the amount of red blood cells that would be available for the luminol to react with. The hematocrit established that the blood used for this experiment consisted of 38% red blood cells, which is within the normal range for female adults.

Four different types of 100% cotton denim were used to give a representation of denim that may be encountered in forensic cases. New denim jeans were purchased and taken directly from the shipping box in the storeroom of the retail establishment to reduce the amount of handling prior to testing. The brand basic concepts, cut 2222, nex style #5130571 was used as the "never-washed, never-worn" sample. Black denim that had previously been machine washed with detergent and worn numerous times was taken from Levi's 550 relaxed fit, #2390ES14. Two different types of blue denim that had been previously been washed and worn numerous times were collected from Express Jeans, RN 55285, style #061451, and Tommy Hilfiger Jeans, #46974. The samples were cut into 3 inch by 3 inch squares, numbered, and assigned a two-digit random number.

Diluted blood was applied to the fabric in the pattern of a two-digit random number (16). Four hundred-microliter aliquots of each dilution were pipetted onto individual denim squares, 200 μL being used for each of the two random digits. One specific dilution was used per denim swatch. In the writing process, the 200 μL of diluted blood for each number was evenly distributed in the pattern of each number through consistent pressure to the pipette. The size of the individual numbers was *c.* 2 inches in height by 1 inch in width. After application of the diluted blood, the denim squares were allowed to dry for *c.* 48 h at room temperature.

The amount of luminol reagent applied to the test squares was standardized. As stated previously, 400 μL of dilutions 1:100, 1:500, and 1:1000 were placed on each of the four types of denim. In this phase of the experiment, the blood dilutions were applied as a single circle in the center of the denim sample of *c.* 5 cm square. The reagent was applied in the form of a fine mist spray from a pump-operated spray bottle. For every pump of the spray bottle, *c.* 0.83 mL of the luminol reagent was dispensed. It was determined that three pumps of the spray bottle (2.5 mL total) was adequate to visualize the circular dilution stains. Consequently, three pumps (*c.* 2.5 mL) of the luminol reagent was used on each of the denim samples in the sensitivity experiment and the results noted accordingly.

The luminol reagent was prepared immediately before the application to the samples. The reagent was prepared according to the formulation suggested by Marie (15): 2 mL 3% H_2O_2 in 50 mL deionized H_2O; 0.05 g luminol (3-Aminophthalhydrazide, 97% Aldrich) in 10 mL of 5% sodium hydroxide. The two solutions are mixed and brought up to 100 mL with deionized H_2O.

Application of the luminol reagent to denim samples was conducted in complete darkness with the samples sprayed one at a time. To eliminate any possible bias in the recording of results, all of the samples were placed in random order, mixing dilution factors and type of denim, and read by an unbiased observer. Samples were sprayed with the luminol solution. Three pumps of the spray bottle handle delivered the 2.5 mL of reagent to each sample.

Once the solution had been sprayed, an unbiased observer was asked to call out whether they saw a number, and if so, what number they witnessed. The observer was told that each sample contained a two-digit number, but nothing else about the experiment. The number called out by the observer was recorded along with sample information before moving onto the next sample.

For this experiment, a result was considered positive if the observer called out the correct number, and negative if the observer did not. All of the numbers were correctly identified by the observer. All dilutions up to 1:1000 were therefore considered to be positive. For each denim class, the greater the dilution, the faster the chemiluminescence faded. When sprayed with luminol, regardless of the dilution, the entire number luminesced, not just the outer edges. When luminescence was noted, it was consistently bright blue in color.

There was no observable difference in the way either of the washed and worn numerous times blue denim samples responded to the luminol. At a dilution of 1:1000 and less, that is, the limit to which the dilutions were carried out in this study, the numbers were easily visible and held the chemiluminescence for greater than 3 sec, with some of the dilutions still faintly luminescent several minutes after the application of luminol.

Compared to blue denim samples, the reaction with black denim was less intense than with blue denim. Though all numbers for all dilutions up to 1:1000 were easily visible when first treated with luminol, the reaction fatigued more quickly at the greater dilutions. All dilutions maintained their chemiluminescence for at least 3 sec, however.

Discussion and Conclusions

In this study, luminol has been found to easily detect bloodstains at dilutions of up to 1:1000 on a variety of denim material. These findings contradict the published account of the FBI Laboratory that luminol is "usually detectable at no more than a 1:100 dilution" on absorbent materials.

This study does support the statement made by Marie (15) that black denim is a more difficult substrate to find blood on when compared to other materials. However, this study does disagree with another statement made by Marie, that "undiluted blood may quench the reaction, although the borders will still glow." In this study, undiluted blood reacted vigorously with the luminol reagent,

and regardless of the dilution, the entire stain showed uniform luminescence, not just the borders.

It is the hope of these authors that this research will renew faith within the forensic community as to the value of the luminol test. Pedantry over the maximum sensitivity of the luminol test is profitless, but concern over the minimum sensitivity is not. The test is sensitive, as others have long maintained, and should occupy a position of respect in the armamentarium of forensic workers.

References

1. Proescher F, Moody AM. Detection of blood by means of chemiluminescence. J Lab Clin Med 1939;24:1183–9.
2. Curtius T, Semper A. Verhalten des 1-athylesters der 3-nitrobenzol-1-2-dicarbonsäure gegen hydrazine. Ber Dtsch Chem Ges 1913;46:1162–71.
3. Albrecht H. Über die chemiluminescenz des aminophthalsäurehydrazids. Z Phys Chem (Leipzig) 1928;136:321–30.
4. Huntress EH, Stanley LN, Parker AS. The preparation of 3-aminophthalhydrazide for use in the demonstration of chemiluminescence. J Am Chem Soc 1934;56:241–2.
5. Geau K, Pfannstiel K. Über 3-aminophthalsäure-hydrazid. J Prakt Chem 1936;146(NF):137–50.
6. Specht W. Die chemiluminescenz des hämins, ein hilfsmittel zur auffindung und erkennung forensich wichtiger blutspuren. Angew Chem 1937;50:155–7.
7. Grodsky M, Wright K, Kirk P. Simplified preliminary blood testing. An improved technique and comparative study of methods. J Crim Law Criminol Police Sci 1951;42:95–104.
8. Castellanó A, Alvarez M, Verdú F. Accuracy, reliability, and safety of luminol in bloodstain investigation. Can Soc Forensic Sci J 2002;35(3):112–23.
9. Weber K. Die anwendung der chemiluminescenz des luminols in der gerichtlichen medizin und toxicologie. I. Der nachweis von blutspuren. Dtsch Z Gesmte Gerichtl Med 1966;57:410–23.
10. Thornton JI, Rios FG, Guarino K, Cashman PJ. Enchancement of the luminol reaction by means of light amplification. J Forensic Sci 1986;31:254–6.
11. Tobe S, Watson N, Daeid N. Evaluation of six presumptive tests for blood, their specificity, sensitivity, and effect on high molecular-weight DNA. J Forensic Sci 2007;52(1):102–9.
12. Bujan V. De quelques changements dans le sang pendant qu'il vieillet et se desseche. Acta Med Leg 1948;1:867–84.
13. Budowle B, Leggitt JL, Defenbaugh DA, Keys KM, Malkiewicz SF. The presumptive reagent fluorescein for detection of dilute bloodstains and subsequent STR typing of recovered DNA. J Forensic Sci 2000;45(5):1090–2.
14. Frégeau CJ, Germain O, Fourney RM. Fingerprint enhancement revisited and effects of blood enhancement chemicals on subsequent profiler plus fluorescent short tandem repeat DNA analysis of fresh and aged bloody fingerprints. J Forensic Sci 2000;45(2):354–80.
15. Marie C. Presumptive testing and enhancement of blood. In: Bevel T, Gardner R, editors. Bloodstain pattern analysis, 3rd edn. Boca Raton, FL: Taylor & Francis, 2008;275–6.
16. Table of Random Numbers. In: Diem K, editor. Documenta Geigy: scientific tables. Ardsley, NY: Geigy Pharmaceuticals, 1962;131.

Additional information and reprint requests:
John Thornton, D.Crim.
Napa County Sheriff's Department
1535 Airport Blvd.
Napa, CA 94558.
E-mail: jthornto@co.napa.ca.us

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01432.x
Available online at: interscience.wiley.com

TECHNICAL NOTE
CRIMINALISTICS

Thatsanee Thonglon,[1] *B.Sc. and Nopadol Chaikum,*[2] *Ph.D.*

Magnetic Fingerprint Powder from a Mineral Indigenous to Thailand

ABSTRACT: A study was conducted to investigate whether natural magnetite (Fe_3O_4), which is an abundant mineral in Thailand, could be used as a magnetic powder in the detection of latent fingerprints. Because of the presence of impurities, powdered magnetite is only weakly attracted by a magnet and cannot be used as a magnetic fingerprint powder by itself. Mixing a small amount of magnetite powder with nickel powder greatly enhances the magnetic attraction. A mixture of magnetite powder and nickel powder in a mass ratio of approximately 1:100 was found to be suitable for use as a magnetic fingerprint powder. Fingerprints developed using the magnetite/nickel mixture on nonporous surfaces were found to exhibit good adherence and clarity. Using an automated fingerprint identification system, the number of minutiae detected in fingerprints developed by using the prepared powder on nonporous surfaces was found to be comparable to those detected in fingerprints developed by using a commercial black magnetic powder. The cost is lowered by more than 60%.

KEYWORDS: forensic science, magnetic powder, fingerprint powder, magnetite, nickel powder, mixed magnetic powder, fingerprint minutiae

Magnetic fingerprint powders are ferromagnetic powders that are applied on latent fingerprints by using a magnetic applicator so that only the powders touch the fingerprint ridges. Iron particles and iron oxide are basic materials of magnetic powders (1). The magnetic particles form the brush and the fine nonmagnetic particles adhere to the fingerprint ridges. The powders are used on smooth surfaces, such as leather where the excess powder can be removed by the applicator. Some magnetic powders have successfully been used to recover latent fingerprints on skin surfaces of living subjects and dead bodies (2). Magnetic flake powders have been manufactured by milling of carbonyl iron and austenitic stainless steel powders and found to be effective for use on both smooth and rough surfaces (3).

Fingerprint powders are important in forensic work especially in crime scene investigation. However, fingerprint powders, especially magnetic powders, are expensive in a country like Thailand, and they must be imported. This investigation is an attempt to prepare a magnetic powder from the mineral magnetite (Fe_3O_4) from an indigenous source and to compare its performance on some smooth, nonporous surfaces with that of a commercially available black magnetic powder.

Materials and Methods

The magnetite was obtained from the Province of Saraburi in central Thailand. The commercial black magnetic powder was purchased from SPEX Forensics (JY Inc., Edison, NJ), and the nickel powder was purchased from S.D. Fine-Chem Limited.

[1]Faculty of Science, Forensic Science Graduate Programme, Mahidol University, Bangkok, Thailand.
[2]Chemistry Department, Faculty of Science, Mahidol University, Bangkok, Thailand.

Received 12 June 2009; and in revised form 11 July 2009; accepted 18 July 2009.

Grinding of the mineral was performed on a Vibro-type ball milling machine (Retsch, Germany). The scanning electron microscope (SEM) image was obtained from an S-2500 SEM (Hitachi, Japan), the X-ray fluorescence (XRF) analysis was carried out on an S4 Explorer X-Ray Fluorescence Spectrometer (Bruker, Germany). The particle size was examined on a Mastersizer 2000 Particle Size Analyzer (Malvern Instruments, U.K.).

The mineral sample was hammered to small pieces and mechanically ground in ethanol for 50 h, and after oven drying at 60°C overnight, the particle size distribution was determined.

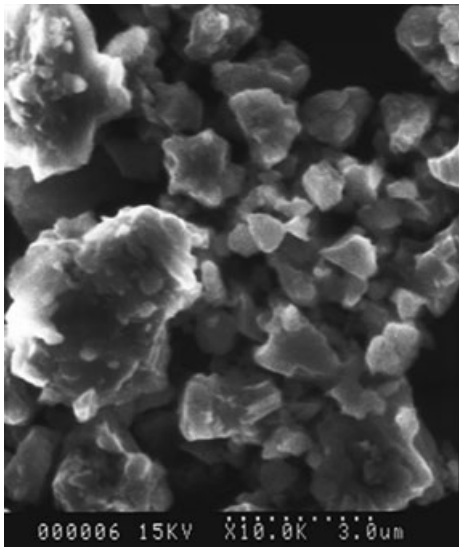


FIG. 1—*SEM image of the powdered magnetite.*

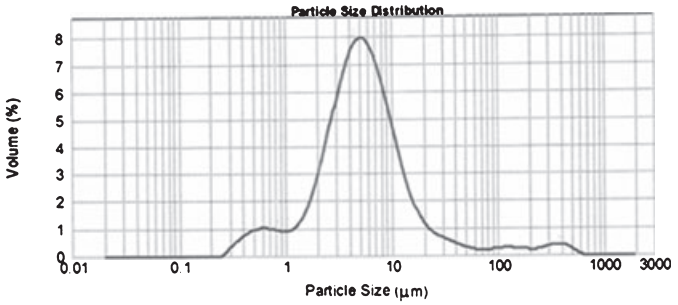


FIG. 2—*Particle size distribution of the powdered magnetite.*

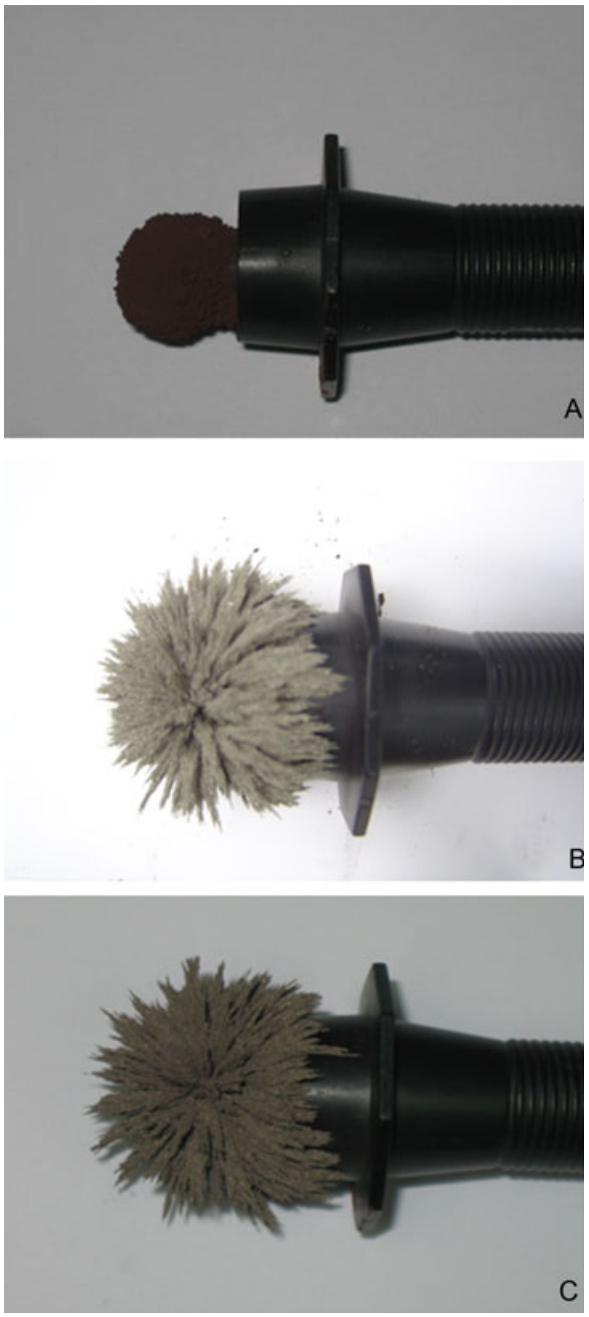


FIG. 3—*(A) Weak attraction between the powdered magnetite and the magnetic applicator. (B) Strong attraction between the Ni powder and the magnetic applicator. (C) Strong attraction between the prepared powder and the magnetic applicator.*

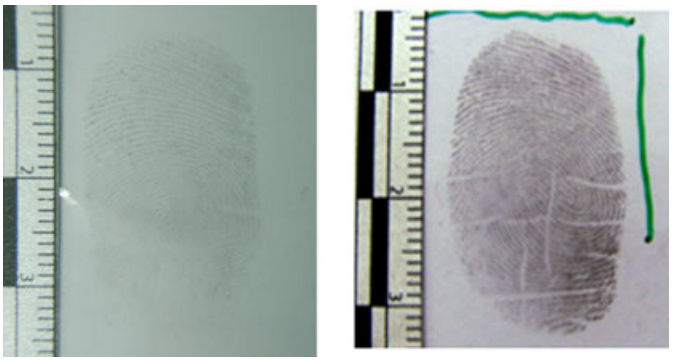


FIG. 4—*Fingerprint on a plastic bag developed by using (A) the Ni powder and (B) the prepared powder.*

Fingerprints were deposited on a white plastic bag, a white ceramic tile, a glass plate, and an aluminum plate. Identical sets of fingerprints were obtained from the same donor who rubbed the right thumb on the nose zone each time before pressing it on the surfaces. The fingerprints were left at an ambient temperature of 25°C for 1, 3, 5, 7, 14, 21, and 28 days. Each experiment was repeated five times.

The prints were dusted with the prepared powder, and another identical set of prints were separately dusted with the commercial black powder using a magnetic applicator. Each print was then tape-lifted and transferred to a fingerprint collection card. The lifted prints were examined on an automated fingerprint identification system (AFIS) and the numbers of minutiae of the prints compared.

Results and Discussion

The XRF analysis of the magnetite showed that the iron content, expressed as Fe_2O_3, was 82.8%. A small amount of aluminum (2.02% as Al_2O_3) was also present.

An SEM image of the powdered mineral is shown in Fig. 1. The particle shapes show some irregularities and jagged edges. Examination of the particle size distribution (Fig. 2) shows that the average particle size of the powdered magnetite was approximately 5.1 µm, while those of the nickel powder and the commercial magnetic powder were found to be about 119.7 and 82.3 µm, respectively. Smaller particles adhere more effectively to fingerprint residue, and rounded particles of *c.* 1 µm in diameter have been most often used in fingerprint powder formulations (4). Fine aluminum, zinc, and iron powders with particulate diameters in the 1–50 µm range, mixed with stearic acid powder, have also been used for fingerprint detection (5,6). Nevertheless, in this investigation, the obtained particle size of magnetite was found to be suitable.

The magnetic property of the powdered magnetite was tested by using a magnetic applicator, and the powder was found to be only weakly attracted to the applicator (Fig. 3*A*), while the nickel powder was strongly attracted (Fig. 3*B*). To increase the magnetic attraction, the powdered magnetite was mixed with the nickel powder on a roller mixer, and the attraction was found to be greatly enhanced (Fig. 3*C*). In fact, the nickel powder itself could be used as a magnetic fingerprint powder, but the developed print was found to be too faint, especially on a light background (Fig. 4*A*). The mass ratios of the magnetite powder to the nickel powder that were found to be suitable for fingerprint development were between 5:100 and 1:100. The mass ratio chosen in this work was 1:100 because fingerprints developed by using this mixture gave the best details with low background (Fig. 4*B*).

Figure 5 shows the minutiae detected by the AFIS for 28-day-old fingerprints developed by using the prepared powder and the commercial black powder on the nonporous surfaces. The developed prints in both cases were visually similar. The numbers of minutiae detected for each pair of fingerprints are comparably high, showing that the quality of the prints is high enough for identification purposes.

Figure 6 shows the variation of the number of minutiae with the age of the fingerprint for both powders. Evidently, as far as the number of minutiae is concerned, the prepared powder performed comparably with the commercial product, although the number of minutiae fluctuated somewhat when the prints were older. A naturally occurring material is inevitably inhomogeneous to some extent and contains a number of impurities. Apparently, these impurities also bound themselves to the moisture and/or the oily components in the fingerprint ridges. These tended to evaporate as the fingerprint aged and the binding became less effective. Nevertheless, the number of minutiae remained fairly high, so that the developed prints remained useful for identification purposes. The mixed powder that had been stored in a desiccator still worked perfectly after 4 months.

The nickel costs approximately 15.6 USD per 100 g, while magnetite costs less than 0.003 USD per 100 g (unground). The commercial magnetic powder costs approximately 52 USD per 100 g. Thus, the prepared powder costs less than one-third of the commercial product.

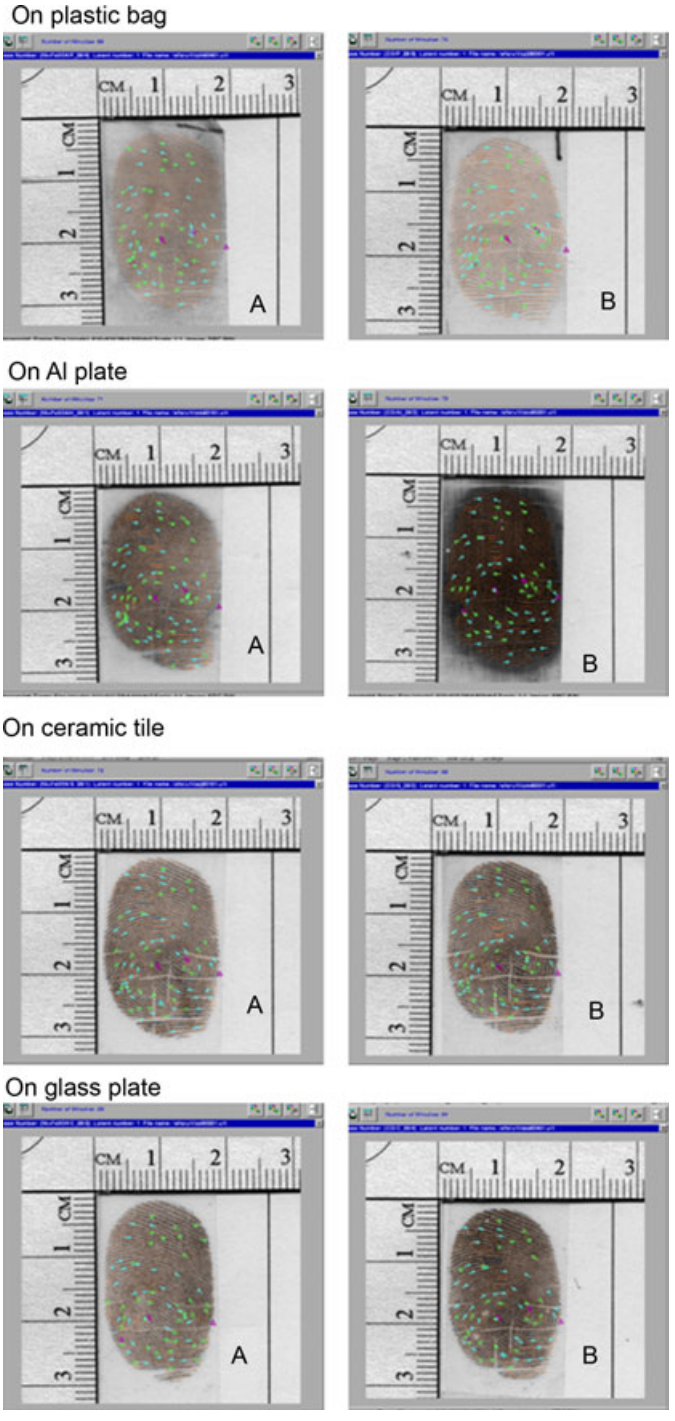


FIG. 5—*Minutiae detected by AFIS from fingerprints developed by using (A) the prepared powder and (B) the commercial magnetic powder.*

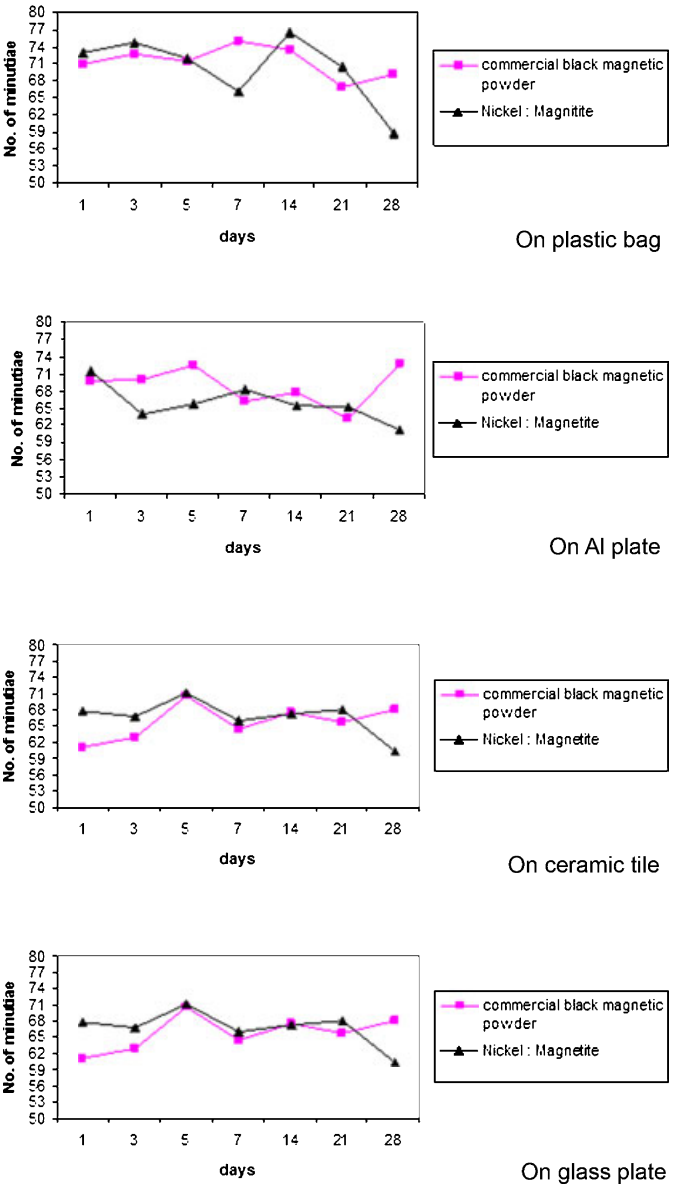


FIG. 6—*Variation of the number of minutiae on 28-day-old fingerprints.*

Conclusion

A magnetic fingerprint powder has been prepared by mixing powdered natural magnetite with nickel powder in a ratio of approximately 1:100. Dusting fingerprints with the mixed powder gave developed fingerprints of quality comparable to those obtained by using a commercial black magnetic powder, judging from the number of minutiae detected by the AFIS. The cost of the prepared powder is less than one-third of that of the commercial magnetic powder. Nevertheless, it should be borne in mind that nickel has been listed as a human carcinogen (7), and the use of a mask is recommended.

Acknowledgments

The use of instruments was made possible by the generosity of the Faculty of Science, Mahidol University. Thanks are due to A. Srisuthipruek for XRF analysis. Thanks are also extended to the Central Institute of Forensic Science, Ministry of Justice for the use of the AFIS.

References

1. Lee HC, Gaensslen RE. Methods of latent fingerprint development. In: Lee HC, Gaensslen RE, editors. Advances in fingerprint technology. New York, NY: CRC Press, 2001;120–5.
2. Trapecar M, Balazic J. Fingerprint recovery from human skin surfaces. Sci Justice 2007;47:136–40.
3. James JD, Pounds CA, Wilshire B. Magnetic flake powders for fingerprint development. J Forensic Sci 1993;38:391–401.
4. Wilshire B. Advances in fingerprint detection. Endeavour 1996;208:12–5.
5. James JD, Pounds CA, Wilshire B. Flake metal powders for revealing latent fingerprints. J Forensic Sci 1991;36:1368–75.
6. James JD, Pounds CA, Wilshire B. Production and characterization of flake metal powders for fingerprint detection. Powder Metall 1991;34:39–43.
7. http://www.lenntech.com/Periodic-chart-elements/Ni-en.ht (accessed on January 12, 2009).

Additional information—reprints not available from author:
Nopadol Chaikum, Ph.D.
Chemistry Department
Faculty of Science, Mahidol University
Rama 6 Road
Bangkok 10400
Thailand
E-mail: scnck@mahidol.ac.th

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01412.x
Available online at: interscience.wiley.com

TECHNICAL NOTE

DIGITAL & MULTIMEDIA SCIENCES

Bart Hoogeboom,[1] *M.Sc. and Ivo Alberink,*[1] *Ph.D.*

Measurement Uncertainty When Estimating the Velocity of an Allegedly Speeding Vehicle from Images

ABSTRACT: Sometimes the question arises whether it is possible to estimate the velocity of a speeding car recorded by closed circuit television cameras. By estimating the travelled distance of the car between two images and the time elapsed, estimation of the velocity is rather straightforward. However, to quantify the corresponding measurement uncertainty, the data analysis becomes more involved. The article describes two approaches as to how to derive the measurement uncertainty. In the first method, distance and timing are estimated separately, and the two uncertainties are combined to derive the measurement uncertainty for the velocity. For this, a frequentist and a Bayesian approach are described. In the second method, the measurement uncertainty for the speed is derived directly using validation recordings of a car driving by at known speed. The choice which method to use depends mainly on the length of the path that the car has travelled.

KEYWORDS: forensic science, digital images, photogrammetry, CCTV images, validation, statistics

The following is a case study on performing a statistically sound analysis on measurement uncertainty when estimating the velocity of an allegedly speeding vehicle from closed circuit television (CCTV) images. From a more general viewing point, it is an illustration on how to quantify measurement uncertainty, as required by the ISO/IEC 17025 standard, in situations where it is not self-explanatory (how) to perform validation experiments and where sample sizes are typically small.

The case can be described as follows. The questioned vehicle is driving through the center of a city at night time and at some moment in time has a collision with another car. Prior to the accident, it is recorded by several CCTV cameras during its journey through the city, and the question arises how fast the vehicle was driving before the collision.

Two moments in time were isolated to estimate the speed of the car. At the first, the car was recorded by two cameras belonging to the same recording system, at the second by a single camera. The two velocity estimations are performed separately, using two different approaches. In the first estimation procedure, footage was available of two cameras connected to the same system, around 600 m apart. Time indications on the cameras show the time interval it took the questioned car to drive from camera 1 to 2, about 28 sec. In the second estimation, the camera footage consisted of two images of the questioned vehicle, taken by the same camera within a limited time interval of *c.* 0.5 sec. The two velocity estimations are treated as two separate cases.

Both velocity estimations were performed by means of a validation experiment, that is, by returning to the scene of crime and measuring known objects in front of the same cameras. The reason for this is twofold:

- The need to avoid systematic errors in the data analysis,
- The need to obtain reliable error readings.

A good way of maintaining the link between case and reality is by gathering validation readings and performing a sound data analysis. For both velocity estimations, we describe the analysis used to obtain reliable confidence intervals. For this, a number of approaches are used to evaluate data when sample sizes are small, based on a result employed earlier in (1) and (2) on body height estimations in images.

Materials and Methods

Situation #1

In the first velocity estimation, according to the street map and the timer of the recording system, the questioned car travelled around 600 m in about 28.23 sec, thus with an estimated mean velocity of *c.* 77 km/h. The car was recorded by two different cameras, connected to the same recording system. The questioned images of the car are shown in Fig. 1. The street between the two camera locations is mostly a straight main road with a number of crossings without traffic lights.

For both street map and timing system, it is unclear how accurate they are. Therefore, an experiment was performed to validate both the estimation of travelled distance and the reported time difference between the questioned images.

To validate the travelled distance, a car of the same type and year as the questioned one was positioned repeatedly by three investigators, at the locations of the questioned images. Here, first, each investigator restored the camera orientation as accurately as possible. The investigators were present in the surveillance camera

[1]Netherlands Forensic Institute, PO Box 24044, 2490 AA, The Hague, The Netherlands.

Received 24 Mar. 2009; and in revised form 3 July 2009; accepted 12 July 2009.

control room, and used overlays of the questioned images on the live camera images. After this, for both locations, the investigators guided a policeman driving the vehicle to the location of the questioned vehicle in the images, again as accurately as possible. The thus positioned cars were measured using land measuring equipment by the police, and for each investigator, a distance resulted between the two locations.

To quantify the distance between the two locations in a confidence band around some point estimate, two assumptions have to be made:

- The readings of the measuring equipment are correct, and
- The position process of the car is accurate, but for some random noise.

The amount of information in the image of the road in the direct vicinity of the car determines whether this second assumption is correct.

To investigate the time difference reported by the camera system, the following was performed. During the reconstruction, two laptops were taken to the questioned cameras with a running timer on screen, as shown in Fig. 2. We refer to the cameras and timers by the numbers 1 and 2.

Timers 1 and 2 were first positioned in front of camera 1. This makes it possible to synchronize the two timers. After this, timer 2 was transported to camera 2, where it was left running for *c*. 10 min. Then, it was transported back to camera 1. In this way, two elements were validated:

- The timing system of the laptop timers, and
- The timing system of the CCTV system.

Validation of the timers was performed by inspecting whether after re-transportation of timer 2, the time difference between timers 1 and 2 was still the same. CCTV times were validated by inspecting 21 disjoint time intervals of *c*. 28 sec. For every interval, the CCTV time difference coming from the time stamps given by the viewer software used to view the video images (estimated time interval) was compared to that of timers 1 and 2 (real time interval). This leads to a histogram of error readings, and a confidence interval can be determined on the error of the original reported time difference of 28.23 sec. Combination of the two confidence bands results in a confidence interval for the mean velocity of the questioned vehicle on the images, as explained in the section Data Analysis.

FIG. 1—*Images of the car recorded by two different cameras.*

FIG. 2—*Recorded image from the two laptops/timers by camera 1.*

Situation #2

In the second velocity estimation, the images of the questioned vehicle were taken by the same camera within *c*. 0.5 sec. The two consecutive images of the car from this camera are shown in Fig. 3.

Here, the questioned vehicle only travelled a limited path. Estimation of the travelled distance would hence probably have a much higher relative measurement uncertainty than on the *c*. 600 m in situation #1. Because of this, validation of the timing and the travelled distance was not treated separately, but validation of velocity was performed directly. To do this, reconstruction images were made of a similar car (same type and year) driving at fixed, known speeds.

The questioned camera recorded test drives of the car driving at a constant speed along the view of the camera, the velocity being recorded by a laser gun. Some 20 test drives were carried out, at different velocities, to obtain a sufficient collection of validation recordings. Only half of the number of test drives could be used for the investigation, because 10 of the test drives yielded only one full image of the car, in which case they were unusable.

Measurements on the test recordings and the questioned recording were performed using a 3D model of the crime scene. Using corresponding points of the questioned images and the 3D model, a virtual camera was derived, looking at the computer model from the same perspective as the real camera. For both images of the car in a sequence, a 3D model of the car was positioned over the car in the image. In this way, the distance between the two positions of the car was estimated. Together with the corresponding reported time interval between the two images, an estimation of the speed would result. For the test recordings, this estimation was compared to the actual velocity (as recorded by the laser gun). In this way, the systematic error and random variation on speed estimations for the questioned car were determined, which led to confidence intervals. The Appendix on data analysis describes how these confidence intervals can be derived. Furthermore, this Appendix describes the statistical analysis which is used in the rest of this technical note.

Results

Situation #1

The measured distances x_1, x_2, and x_3 by the three investigators equal 629.0, 629.2, and 630.0 m. The mean distance between these three measurements is $\bar{x} = 629.4$ m, with a standard deviation $s_X = 0.53$ m. The standard deviation can be used to calculate the borders of a 95% confidence interval, using the test statistic *U*. Assuming

FIG. 3—*The two images of the car for which a velocity estimation was required.*

normality of the sample, which is reasonable, a $\sqrt{0.95} \times 100\%$ confidence interval for the real distance μ is given by

$$\bar{x} \pm q_U \times (s_X/\sqrt{3}) = 629.4 \pm 6.16 \times (0.53/\sqrt{3})\text{m} = 629.4 \pm 1.8 \text{ m}$$

For the timers on the laptops, as a first validation check, they were checked by recording the time for 1 min with a handy-cam recording at 50 fields/sec. This showed no problems. When the timers on the laptops were started, the time difference between the timers was 0.03 sec. At the end of the timer recordings, after *c*. 15 min, the time difference was 0.00 sec. For the 28.23 sec interval in the current case, this would imply a shift of *c*. 10^{-3} sec in time difference, which for our purposes is negligible.

In Fig. 4, the observed differences $\delta_1,\ldots, \delta_{21}$ between the intervals from the timer and the camera system are presented in a histogram.

The mean difference of these 21 intervals is $\bar{\delta} = 0.017$ sec, with a standard deviation of $s_X = 0.061$ sec. The Shapiro–Wilks test on normality of the sample does not reject the hypothesis of normality ($p = 0.085$). The standard deviation determines a confidence band for the questioned recording. A $\sqrt{0.95} \times 100\%$ confidence interval for the difference Δ between reported and actual time interval on the questioned images is given by

$$\bar{\delta} \pm q_R\sqrt{1 + 1/21}\, s_\Delta = 0.017 \pm 2.42 \times 1.024 \times 0.061\text{sec} = 0.017 \pm 0.015 \text{ sec}$$

In the Appendix, two approaches are described to combine the confidence intervals for the distance and the time to a confidence interval for the speed. The first, straightforward, approach leads to a 95% confidence interval for the mean velocity *v* of the car of

$$\frac{629.4 - 1.8m}{28.227 + 0.032s} \le v \le \frac{629.4 + 1.8m}{28.227 - 0.002s},$$

or

$$22.2 \text{ m/sec} \le v \le 22.4 \text{ m/sec},$$

or

$$79.9 \text{ km/h} \le v \le 80.5 \text{ km/h}.$$

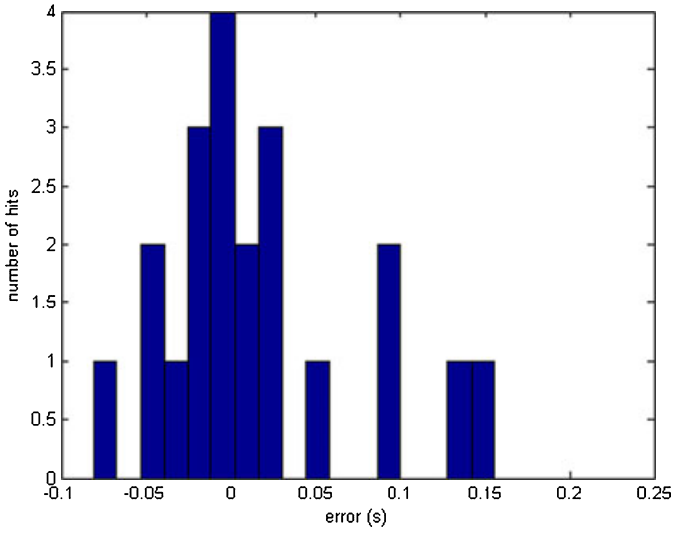


FIG. 4—*Histogram of the differences between the time interval readings from the timer and the time given by the camera system.*

Next we follow the Bayesian approach described. Conditional on $x_1,\ldots,x_3$ and $\delta_1,\ldots,\delta_{21}$ and t_{observed}, we have

$$\frac{\mu - \bar{x}}{s_X/\sqrt{3}} \sim t_2 \quad \text{and} \quad \frac{\Delta - \bar{\delta}}{\sqrt{1 + 1/21}\, s_\Delta} \sim t_{20},$$

where $\sim t_m$ means that the statistic to the left has the so-called Student's *t* distribution with *m* degrees of freedom. Furthermore, μ and Δ are statistically independent, which determines the probability distribution of the velocity

$$V = \mu/(t_{\text{observed}} + \Delta).$$

In fact, we are only interested in a 95% credible interval for *v*, which we obtain by taking two large samples (of size 100,000) from the t_2 and the t_{20} distribution, simulating 100,000 outcomes for the combination (μ,Δ) and evaluating the corresponding values of *v*. An example of a histogram of the values obtained is given in Fig. 5.

It turns out that

$$\text{P}(79.9 \text{ km/h} \le V \le 80.5 \text{ km/h}) = 0.95,$$

which is robust under repetition of the simulation. All in all, we obtain the same interval as in the frequentist approach.

Situation #2

As explained, we have deviations between real and measured speed, which we will refer to as $\Delta_j = \Delta V_j$, for $j = 1,\ldots,10$. Based on this, we want to predict the velocity deviation $\Delta = \Delta V$ for the questioned images. The speed derived from the images and the speed recorded by the laser gun is given in Table 1.

The mean difference between the velocity derived by the laser gun or from the images is $\bar{\delta} = 0.18$ m/sec, with standard deviation $s_\Delta = 0.67$ m/sec. The recording circumstances between the validation recordings and the questioned recording are assumed identical. Under the conditions as stated, following the Appendix, the statistic

$$R = \frac{\Delta - \bar{\Delta}}{\sqrt{1 + 1/10}\, S_\Delta}$$

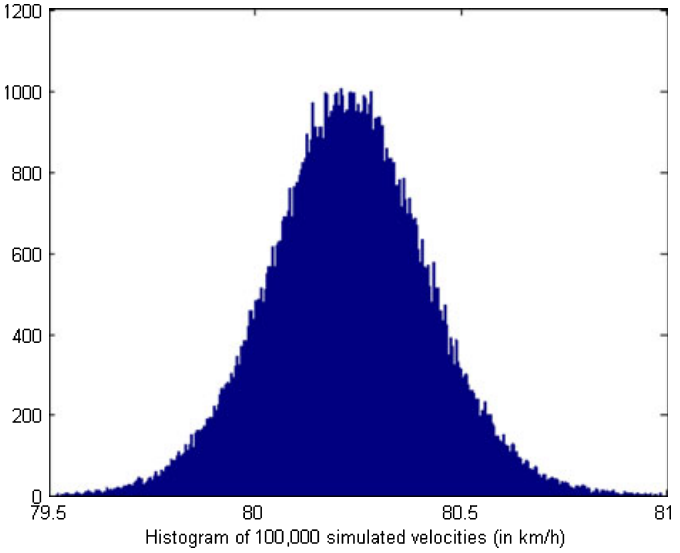


FIG. 5—*Histogram of 100,000 simulated velocities, following the Bayesian model.*

TABLE 1—*Velocities derived by the laser gun compared to those obtained from images.*

Type of ride	Speed Recorded by the Laser Gun (m/sec)	Speed Derived From the Images (m/sec)	Difference (m/sec)
Validation	16.1	16.5	−0.4
Validation	18.6	17.9	0.7
Validation	19.2	19.0	0.2
Validation	15.3	15.2	0.1
Validation	19.2	18.1	1.1
Validation	16.4	16.1	0.3
Validation	16.1	15.6	0.5
Validation	15.8	14.9	0.9
Validation	19.2	19.9	−0.7
Validation	20.3	21.2	−0.9
Questioned	–	19.5	–

has the Student's t distribution with 10−1 = 9 degrees of freedom. On the basis of this, a 95% confidence interval for ΔV is obtained. The mean difference 0.18 m/sec is used as a correction on the derived speed 19.5 m/sec on the questioned car. The standard deviation $s_\Delta = 0.67$ m/sec determines a confidence band for the questioned recording. In this way, a 95% confidence interval is obtained of

$$19.5 + 0.18 \pm \xi_{10-1,0.975} \times \sqrt{(1 + 1/10)} \times 0.67 \text{ m/s} = 19.7 \pm 1.6 \text{ m/s}.$$

Here by $\xi_{10-1,0.975} = 2.26$ we mean the 97.5% quantile corresponding to the Student's t distribution with 10−1 = 9 degrees of freedom. That is, the 95% confidence interval for the velocity v of the car at the location shown in Fig. 4 is:

$$65 \text{ km/h} \leq v \leq 77 \text{ km/h}.$$

Conclusion

When estimating the velocity of a speeding vehicle from video footage, an estimation of the measurement uncertainty is needed. In this article, two approaches are described, both leading to a confidence interval for the velocity. The similarity between these approaches is that they both use validation recordings of known events. Validation recordings are necessary to obtain confidence intervals around estimated speeds. If the original CCTV system is no longer available, the suggested method cannot be used.

In the first approach, travelled distance and time interval are estimated separately, by using validation recordings of a positioned car and validation recordings of a stopwatch. The separate measurement uncertainties are combined to derive the measurement uncertainty for the velocity. To this, a combination of the two estimation procedures has to take place, which is described in both a so-called frequentist and a Bayesian approach. In the second method, the measurement uncertainty for the speed is derived directly using validation recordings of a car driving by at a known speed. Here, a result employed earlier in (1) and (2) is used.

In the case study, high accuracy was achieved (±0.5%) when the speed was determined over a long distance and time. A drawback of this is of course that because the speed is averaged over a long distance, it may be a gross underestimation of the top speed actually driven along that path.

As stated in the Introduction, the above is an illustration of how to quantify measurement uncertainty, as required by the ISO/IEC 17025 standard, in situations where it is not self-explanatory (how) to perform validation experiments and where sample sizes are typically small. Here, the result presented as Theorem A1 in (1) usually is usable. Moreover, the above illustrates how to go about when combining confidence regions for different quantities.

Acknowledgments

The authors thank their colleagues Arnout Ruifrok and Reinoud Stoel for their thorough reviews of the article.

References

1. Alberink I, Bolck A Obtaining confidence intervals and likelihood ratios for body height estimations in images. Forensic Sci Int 2008; 177(2–3):228–237.
2. Hoogeboom B, Alberink I, Goos M. Body height measurements in images. J Forensic Sci 2009;54(6):1365–1375.
3. Gelman A, Carlin JB, Stern HS. Bayesian data analysis, 2nd edn. Boca Raton, FL: Chapman & Hall/CRC, 2004.
4. Aitken C, Lucy D. Estimation of the quantity of a drug in a consignment from measurements on a sample. J Forensic Sci 2002;47:968–975.

Additional information and reprint requests:
Bart Hoogeboom, M.Sc.
Laan van Ypenburg 6
2497 GB The Hague
The Netherlands
E-mail: hoogeboom@holmes.nl

Appendix: Data Analysis

The data analysis addresses two issues. For situation #1, the question is how to combine the measured distances and timer deviations into a confidence interval for the velocity of the car. For situation #2, the question is how to obtain a confidence interval for the velocity of the car based on observed differences between calculated and real (laser gun) velocities.

From here on, we follow statistical tradition to use capital letters to denote random variables, and small ones for outcomes of the latter.

We start with (easier) situation #2. Here we have deviations between real and measured speed, which we will refer to as $\Delta_j = \Delta V_j$, for $j = 1,\ldots,10$. Based on this, we want to predict the velocity deviation $\Delta = \Delta V$ for the questioned images. We can then use a statistical result presented in (1) as Theorem A1, which tells that if the mean and the variance of the sample are denoted by $\bar{\Delta}$ and S_Δ, under normality assumptions on the sample, the statistic

$$R = \frac{\Delta - \bar{\Delta}}{\sqrt{1+1/10}\, S_\Delta}$$

has the Student's t distribution with 10−1 = 9 degrees of freedom. On the basis of this confidence intervals for Δ are readily obtained.

For situation #1, the positioning of the car was done by three operators, so we have three distance measurements $X_1,\ldots,X_3$. The timing was investigated using 21 intervals of about 28 sec, which yielded deviations between system and real time intervals $\Delta_j := \Delta t_j$, for $j = 1,\ldots,21$. Finally there is the observed questioned time interval t_{observed} from the images of the questioned recording. For distance and time individually it is not complicated to obtain confidence regions. Indeed, as to the travelled distance, a classical result states that if the real travelled distance is μ, and assuming normality, the statistic

$$U = (\bar{X} - \mu)/(S_X/\sqrt{3})$$

has the Student's t distribution with $3-1 = 2$ degrees of freedom. For the time interval, let the mean and the variance of the deviations $\Delta_j = \Delta t_j$ again be denoted by $\bar{\Delta}$ and S_Δ. Now for the real questioned time interval

$$T_{\text{real}} = T_{\text{observed}} + \Delta$$

the result presented above applies, that is, the test statistic

$$R = \frac{T_{\text{real}} - T_{\text{observed}} - \bar{\Delta}}{\sqrt{1 + 1/21}\, S_\Delta}$$

has the Student's t distribution with $21-1 = 20$ degrees of freedom. Given T_{observed}, $\bar{\Delta}$ and S_Δ this leads to a confidence interval for T_{real}. We describe a frequentist and a Bayesian approach in which to combine the two.

The first approach is straightforward: using the two statistics described, and the fact that it is natural to assume that they are independent, one can choose quantiles q_U and q_R such that

$$\mathrm{P}(|U| < q_U \text{ and } |R| < q_R) = 0.95.$$

In fact this is, for example, achieved by taking

$$q_U = 6.16 \text{ and } q_R = 2.42.$$

In this case we have that $\mathrm{P}(|U| < q_U) = \mathrm{P}(|R| < q_R) = \sqrt{0.95}$, and because of independence then $\mathrm{P}(|U| < q_U \text{ and } |R| < q_R) = \mathrm{P}(|U| < q_U)\ \mathrm{P}(|R| < q_R) = 0.95$.

Given the outcomes for the X_j and Δ_j this leads to a 95% confidence region for the combination (μ,Δ), which can be translated in an interval for the mean velocity of ≥95% confidence. Note that this may mean that the reported interval is wider than necessary for 95% confidence. Of course we may use alternative quantiles q_U and q_R here, but they have to be fixed prior to the experiment, so we cannot optimize our estimation of v in this respect.

The Bayesian approach is more involved. Here the travelled distance μ and the questioned time deviation Δ are considered as random variables, instead of fixed numbers, with so-called non-informative prior distributions

$$p_1(\mu_1, \sigma_1^2) \propto (\sigma_1^2)^{-1} \text{ and } p_2(\mu_2, \sigma_2^2) \propto (\sigma_2^2)^{-1}.$$

The reader is referred to (3) for the notation used. As a result then the probability distribution of U and R given the outcomes of the samples $x_1,\ldots,x_3$ and $\delta_1,\ldots,\delta_{21}$ and t_{observed} is the Student's t distribution with 2 and 20 degrees of freedom, cf. (3), section 3.2. This fixes the posterior probability distribution for μ and Δ. Since they are independent, the posterior probability distribution of the mean velocity

$$V = \mu/(t_{\text{observed}} + \Delta)$$

of the car is then known in principle, and quantiles can be determined by simulation of the distribution. The reader is referred to (4), in which a similar approach is used.

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01452.x
Available online at: interscience.wiley.com

TECHNICAL NOTE

GENERAL; PSYCHIATRY & BEHAVIORAL SCIENCES

Sean A. Spence,[1] *M.D., F.R.C.Psych.; Alexandra Hope-Urwin,*[1] *R.N., M.Med.Sci.; Sudheer T. Lankappa,*[1] *M.R.C.Psych.; Jean Woodhead*[1]*; Jenny C.L. Burgess,*[1] *B.Med.Sci.; and Alice V. Mackay,*[1] *B.Med.Sci.*

If Brain Scans Really Detected Deception, Who Would Volunteer to be Scanned?

ABSTRACT: Recent neuroimaging studies investigating the neural correlates of deception among healthy people, have raised the possibility that such methods may eventually be applied during legal proceedings. Were this so, who would volunteer to be scanned? We report a "natural experiment" casting some light upon this question. Following broadcast of a television series describing our team's investigative neuroimaging of deception in 2007, we received unsolicited (public) correspondence for 12 months. Using a customized template to examine this material, three independent assessors unanimously rated 30 of an initial 56 communications as unequivocally constituting requests for a "scan" (to demonstrate their author's "innocence"). Compared with the rest, these index communications were more likely to originate from incarcerated males, who were also more likely to engage in further correspondence. Hence, in conclusion, if neuroimaging were to become an acceptable means of demonstrating innocence then incarcerated males may well constitute those volunteering for such investigation.

KEYWORDS: forensic science, functional neuroimaging, functional magnetic resonance imaging, deception, volunteers, male prisoners

Since 2001, more than 20 studies have appeared in the peer-reviewed scientific literature purporting to describe the functional anatomical correlates of deception in the human brain (i.e., the location of neural activity associated with lying) (1). Such studies have utilized brain-imaging techniques such as functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) to detect the location of cerebral neural activity. Although there have been some notable inconsistencies across studies (2), there has nevertheless emerged a recurrent pattern of findings suggesting that at some point in the future functional neuroimaging may be used to detect deception in situations that have significant societal consequence, e.g., legal proceedings. So far, only one published study has examined a "real" forensic problem: the case of a woman convicted of harming a child in her care, who continues to allege a miscarriage of justice (3). However, most published studies (to date) have concerned the neural activity of healthy college students or professionals participating in brain-imaging experiments, which have involved "lying" in rather "low-stake" situations. While the findings of such studies have exhibited a certain consistency (there is usually increased prefrontal cortical activity detectable during attempted deception, compared with truthfulness), they are clearly derived from atypical sectors of society (e.g., those who have undergone further or "higher" education).

Hence, if brain imaging were to be incorporated into legal proceedings (assuming that the relevant technical and ethical problems associated with the process could be adequately resolved) (2), the question would remain as to who might wish to avail themselves of such a technique. Assuming that it would always be ethically preferable to study solely those subjects who *willingly* volunteered for such procedures, which group(s) of people might be expected to come forward? Our own conjecture has been that they would probably comprise those people who espouse the belief that they have been subject to a miscarriage of justice; however, until now, this proposal has been purely speculative (4).

Recently, an opportunity arose to examine this question as a "natural experiment," following the broadcast of a television series in Britain in June 2007. The three programs comprising this series followed the progress of a group of subjects who underwent brain-imaging experiments in our laboratory; experiments involved truths and lies that the participants were invited to tell (while undergoing fMRI), concerning alleged offenses that they may or may not have committed (or been party to). The subjects were all volunteers, recruited by a television production company, and each had a specific (and unique) case to make. The programs depicted the method of brain imaging used and also the analytic techniques that might be applied to the resulting data, to detect truth- and lie-related neural activity. Following the transmission of these films, the principal investigator (SAS) received unsolicited correspondence from members of the public (for a period of 12 months), some of it clearly outlining the correspondent's desire to be scanned, as a way of "proving" their innocence of various alleged offenses. These communications (by letter and email) form the raw data of this report, which sets out to examine the characteristics of their authors (our correspondents), in so far as these might reasonably be deduced from spontaneous communications (of varying length and grammatical construction). Hence, we have utilized an unanticipated opportunity to sample a spontaneously acquired dataset emerging "naturally" in response to a television series.

[1]The University of Sheffield, Academic Clinical Psychiatry, The Longley Centre, Norwood Grange Drive, Sheffield S5 7JT, U.K.

Received 5 May 2009; and in revised form 13 July 2009; accepted 23 July 2009.

Methods

The television series *("Lie Lab")* was broadcast on Channel 4 Television in Britain on the Saturday evenings of 2nd, 9th, and 16th of June 2007. The estimated viewing figures for the first broadcast were 600,000 viewers or 4% of audience share for the hour of transmission (5). We do not know how many people watched the subsequent programs, but Channel 4 is one of the five main terrestrial channels currently broadcasting throughout Britain.

Following transmission of the series, we began to receive letters and emails from the public. Often the manifest content concerned the correspondent's desire to undergo a brain scan, similar to those depicted in the TV series, to prove their innocence of an alleged offense or else to nominate someone they knew to undergo such a procedure. In each case, we replied with a unique letter to the correspondent (acknowledging the specific content of their particular case), but there were always two essential points included in each of our replies:

- The fact that functional brain-imaging evidence (derived from fMRI) would not be currently admissible in a court of law in Britain; and
- That the fMRI technique remains relatively costly to perform and would probably be prohibitively expensive for most private individuals to finance from their own funds.

At no point did we offer to scan any of our correspondents and we have not met with any of them in person. Nevertheless, some correspondents continued to communicate with us: sometimes they sent further details clarifying their case or else expressed the desire to be considered for future studies emerging from our unit. These subsequent communications were also retained on file. In total, we retained all correspondence that arrived within one calendar year of the television series being broadcast.

Subsequently, we (SAS) devised a one-page proforma questionnaire that would be used to extract and record objective information from each communication received; we then used this questionnaire to extract the data reported in the following paragraph. Three assessors, each "blind" to their colleagues' assessments of the same communication, assessed each item of initial correspondence from each correspondent. The three assessors were drawn from a pool of four investigators (AHU, STL, JCLB, and AVM) randomly assigned to each item of (initial) correspondence by another colleague (JW). Hence, while each item of correspondence was independently assessed by three investigators, the allocation of specific assessors varied across the dataset. As a consequence, we have placed greatest emphasis upon those ratings that were scored unanimously by all three assigned assessors (below).

Data were recorded (in SPSS version 14.0; IBM, Chicago, IL) concerning the group of correspondents as a whole (i.e., the "total sample") and subsequently divided between those who appeared to have unequivocally volunteered themselves as subjects for a brain scan ("volunteers") and those for whom this was not the case ("others"; as assessed by our investigators; above).

Following collation of all data, we applied simple descriptive statistics to the "total sample's" correspondence and chi-square analyses when comparing putative "volunteers" with "others."

Results and Discussion

Total Sample

In all, we received initial communications from 56 correspondents. Forty-one correspondents (73%) had sent letters and 15 (27%) emails; 32 communications were "typed" (i.e., the 15 emails plus 17 letters), while 24 were handwritten (letters). An author was identified by name in all 56 communications, being apparently male in 41 cases (73%) and female in 13 (23%; two names appeared ambiguous).

A postal, mail address was supplied by 45 correspondents (80%). Twenty-five communications had been sent from prison addresses (45%; all located within Britain), and none had been sent from hospital addresses.

In 30 cases (54% of the sample), the correspondent was judged by three assessors as unequivocally volunteering for a brain scan, while in 11 (20%) the correspondent appeared to be nominating someone else for such a procedure. The putative subject (either volunteering or being nominated) was men in 35 cases (63%) and women in 7 (13%). (Please note that the sum of putative subjects is greater than the number of correspondents volunteering or nominating others because sometimes a correspondent volunteered themselves while also nominating someone else for a brain scan.)

In 22 communications (39%), there were descriptions of (one or more) offenses for which someone had been convicted. Eleven violent offenses, 11 sexual offenses, 4 financial offenses and 9 offenses involving children were described. (Again, the sum of offenses was greater than the number of communications describing offenses as some of the latter described multiple alleged offenses.)

Thirty-two correspondents (57%) sent follow-up communications within one calendar year of the initial TV series (i.e., before July of 2008).

"*Volunteers*" *Versus* "*Others*"

Comparison of the 30 initial communications from those who volunteered themselves for brain scans with the correspondents categorized as "others" ($n = 26$; above), revealed some statistically significant differences (Table 1). Those communications arising from apparent "volunteers" were statistically more likely to have been sent from prison addresses ($p < 0.01$), to propose the scanning of a male subject ($p < 0.04$), and to be followed up by further correspondence within a year ($p < 0.04$). At the level of statistical trends (i.e., $p < 0.1$), they were also more likely to have been sent by apparently male correspondents and to describe an offense for which someone had been convicted. With respect to those communications arising from the "others" group, these were more likely to contain the nomination of a subject other than the author (to undergo a brain scanning procedure; $p < 0.04$).

Discussion

Our data have been derived from a "natural experiment," which presented itself as an opportunity afforded to us, in the wake of a television series broadcast in Britain in the summer of 2007. As such, these data will comprise correspondence originating from a highly selected sample of individuals (those who had seen one or more of the three TV programs and taken the time and effort to correspond with the principal investigator). Under "normal" circumstances such a highly selected group of subjects might comprise a biased (and, hence, characteristically flawed) dataset; however, in the current instance, we believe that the self-selection by our subjects (correspondents) of themselves is absolutely germane to the central question under investigation: namely, "who would wish to undergo a brain scan if they *really* believed that it might demonstrate the difference between truth and deception, innocence and guilt?" Hence, we should anticipate that subjects volunteering themselves for such a procedure might indeed share

TABLE 1—*Characteristics of communications sent by those correspondents who volunteered for brain scanning and "other" correspondents (total* N = *56).*

Characteristic	"Volunteers" (*N* = 30; 100%)	"Others" (*N* = 26; 100%)
Correspondence is by letter	25 (83%)	16 (62%)
Correspondence is by email	5 (17%)	10 (39%)
Correspondence is typed	16 (53%)	16 (62%)
Correspondence is handwritten	14 (47%)	10 (39%)
Author is identified	30 (100%)	26 (100%)
Author is apparently male	26 (87%)	15 (58%)*
Author is apparently female	4 (13%)	9 (35%)
Postal mail address is supplied	25 (83%)	20 (77%)
Correspondence comes from a prison	19 (63%)	6 (23%)†
Correspondence comes from a hospital	0	0
Correspondent volunteers to be scanned	30 (100%)	0
Correspondent nominates someone else to be scanned	2 (7%)	9 (35%)‡
Proposed scan subject is male	26 (87%)	9 (35%)§
Proposed scan subject is female	4 (13%)	3 (12%)
Correspondence describes an offense for which someone has been convicted	15 (50%)	7 (30%)¶
Correspondence describes a violent offense	7 (23%)	4 (15%)
Correspondence describes a sexual offense	8 (27%)	3 (12%)
Correspondence describes a financial offense	4 (13%)	0
Correspondence describes an offense involving a child	4 (13%)	5 (20%)
The correspondent writes more than once	22 (73%)	10 (38%)**

*$\chi^2 = 2.95$, df = 1, $p < 0.09$ (trend).
†$\chi^2 = 6.76$, df = 1, $p < 0.01$.
‡$\chi^2 = 4.46$, df = 1, $p < 0.04$.
§$\chi^2 = 8.26$, df = 1, $p < 0.01$.
¶$\chi^2 = 2.91$, df = 1, $p < 0.09$ (trend).
**$\chi^2 = 4.50$, df = 1, $p < 0.04$.

certain unusual characteristics, features that might distinguish them from a "normal" or randomly acquired sample of the general population.

Though our sample is small, it is clear that correspondents were predominantly men and that just under half were corresponding from prison addresses within Britain. The latter point is subject to verification because letters sent from prisons all carry identifying inscriptions, including the inmate-correspondent's prison number. Hence, we can be cautiously optimistic that we have accurately identified the names and addresses of each of our incarcerated correspondents.

Nevertheless, we have no means of corroborating the *content* of the correspondence we received; we have not verified whether accounts of putative events and offenses have bases in fact, though, again, this does not nullify the central premise of our investigation: we set out to discover what groups of people might *volunteer* for a brain scan (aimed at diagnosing veracity). Therefore, in so far as we are interested in who might come forward, our correspondents' veracity or accuracy are not necessarily pivotal to whether they might one day constitute the voluntary subjects of such an intervention (fMRI brain scanning).

With respect to the specific characteristics of those 30 correspondents who appeared to be volunteering *themselves* for scanning, we must be circumspect and attempt to avoid reading too much into the data acquired from such a small sample. However, within the context of this spontaneously emerging dataset, and in comparison with our "others" correspondents, our "volunteers" more often wrote from a prison (in nearly two-thirds of cases) describing an offense for which someone had reportedly been convicted (in half of cases; a statistical trend, $p < 0.1$) and they were more likely to send follow-up correspondence (in nearly three-quarters of cases). Furthermore, though their being apparently "male" reached only a trend level of statistical significance (when compared with "others" correspondents), our volunteers were significantly more likely to have proposed that a male subject undergo brain scan investigation. Unsurprisingly, "others" correspondents were significantly more likely to have proposed that another individual undergo such investigation.

There are caveats that are important to acknowledge: not least the self-selected nature of our sample and its relatively small size. While self-selection is not an insurmountable problem, given the nature of our inquiry (above), the small sample size may have exposed us to the risk of type 2 errors (i.e., "false-negatives"). This seems particularly likely with respect to intergroup comparisons where there appear to be numerical differences that have not satisfied standard levels of statistical significance (e.g., the apparent differences in modes of communication, letters versus emails, and the reported nature of the alleged offenses described in accounts provided by "volunteers" and "others"; see Table 1).

There is also a qualitative issue that must be considered: whether volunteering for a scan "really" represents a sincere desire to undergo investigation (to establish innocence) or perhaps constitutes a means to some other end(s)? A sceptical interpretation of our data might be that those who are imprisoned merely volunteered to be scanned as a means of leaving the confines of prison for the duration of investigation (perhaps as a prelude to absconding). Such an interpretation is speculative but it does highlight the need for care when progressing this area of investigative science.

Nevertheless, despite these caveats, our findings still have two specific implications for further investigations in this field, particularly if there is a future move toward the application of functional neuroimaging to the forensic detection of deception and truth, guilt and innocence:

- Given the preponderance of incarcerated men among those volunteering for such investigations, we should propose that future neuroimaging studies of deception should at least attempt to assay those who have been incarcerated (not least because such people may constitute the appropriate "control" groups in any future applications of this technology, within the legal setting);
- Because of the particular life circumstances of those who may volunteer in future (people who are imprisoned, and who may regard themselves as the subjects of miscarriages of justice),

there is an obligation, incumbent upon those who undertake such investigations, to avoid exploitation of potential volunteers' desperation or credulity during any future recruitment process. Hence, we (as a scientific community) must avoid offering false hope to those who may be living in very difficult circumstances, and we must also ensure that consent to investigation is non-coerced and freely given.

Hence, future studies of the neural correlates of deception need to devote more attention to people (especially males) who are or have been incarcerated.

Acknowledgments

We are grateful to those who took the time and effort to correspond with us and we are also grateful to Quickfire Media and Channel 4, who funded the television series concerned.

References

1. Spence SA, Farrow TFD, Herford AE, Wilkinson ID, Zheng Y, Woodruff PWR. The behavioural and functional anatomical correlates of deception in humans. NeuroReport 2001;12:2849–53.
2. Spence SA. Playing Devil's advocate: the case against fMRI lie detection. Legal Criminol Psychol 2008;13:11–25.
3. Spence SA, Kaylor-Hughes CJ, Brook ML, Lankappa ST, Wilkinson ID. 'Munchausen syndrome by proxy' or a 'miscarriage of justice': an initial application of functional neuroimaging to the question of guilt versus innocence. Eur Psychiatry 2008;23:309–14.
4. Spence SA, Kaylor-Hughes CJ. Looking for truth and finding lies: the prospects for a nascent neuroimaging of deception. Neurocase 2008;14: 68–81.
5. http://www.guardian.co.uk/media/2007/jun/04/overnights

Additional information and reprint requests:
Sean A. Spence, M.D., F.R.C.Psych.
Professor of General Adult Psychiatry
Academic Clinical Psychiatry
University of Sheffield
The Longley Centre
Norwood Grange Drive
Sheffield S5 7JT
U.K.
E-mail: s.a.spence@sheffield.ac.uk, or jean.woodhead@sheffield.ac.uk

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01442.x
Available online at: interscience.wiley.com

TECHNICAL NOTE

PATHOLOGY/BIOLOGY

Roger W. Byard,[1,2] *M.D. and Maria Bellis,*[2] *B.Sc.*

The Effect of Decalcifying Solutions on Hemosiderin Staining

ABSTRACT: To determine whether routine decalcification may reduce the amount of stainable iron that is visible on tissue sections, samples of liver and lung tissue with excessive iron stores were placed in three standard decalcifying solutions (i) formic acid [33%], formaldehyde [4%], and NaCl [0.85%]; (ii) formic acid [30%], formaldehyde [4%], and water; and (iii) nitric acid [5%] for 24, 48, 72, and 96 h. After exposure to the decalcifying solutions, the tissues were stained with Perls stain. The slides were examined blind and the intensity of iron staining was scored semiquantitatively from 0 to 3+. The trend in all samples over the course of the experiment (96 h) was for reduction in the intensity of hemosiderin staining. As the amount of stainable hemosiderin in tissues may be significantly altered by decalcification, the absence of hemosiderin in tissues adjacent to a fracture site does not necessarily indicate that the injury was acute.

KEYWORDS: forensic science, iron staining, hemosiderin, Perls, fractures, child abuse, dating of injuries, inflicted injuries

The dating of injuries in a medicolegal context may play an important role in helping to determine when a particular accident or assault may have occurred. This may have significance in deciding who may, or may not, have had an opportunity to cause the injury. For example, hemosiderin derived from red cell breakdown has been used as an indicator that an injury was sustained some time (days) before death.

Unfortunately, attempting to interpret findings in a specific case based on generalizations from the literature, or even from individual experience, may be difficult, as the changes that occur in injuries over time are often idiosyncratic and do not always follow precise chronological sequences. An example of this concerns the great variation in the appearances of bruises that occurs among individuals, and even within the same individuals at different anatomical sites, over the same period of time (1,2). In addition, laboratory processing may also on occasion affect the microscopic appearance of tissues.

Given the potential significance of the presence or absence of hemosiderin in sections from fractures in infants and children, the following study was undertaken to determine whether staining of tissues for hemosiderin, resulting from red blood cell break down, may be affected by the process of decalcification.

Materials and Methods

Control tissues are routinely processed in histopathology laboratories to ensure the validity of staining reactions when special stains are being performed. This process is undertaken at Forensic Science SA in accordance with national (NATA) laboratory guidelines. Samples of liver and lung tissue with excessive iron stores from six cases (three liver, three lung) that had been stored in 10% buffered formalin and used for controls were divided into 30 cassette-sized portions (15 liver, 15 lung). Bone was not used in this study as sections had to be able to be cut before the tissues were exposed to decalcification to provide a baseline for hemosiderin staining.

A sample from each case was routinely stained with Perls stain for hemosiderin according to standard protocols prior to exposure to decalcifying agents ($n = 6$). The remainder of the samples were then placed in three standard decalcifying solutions (i) formic acid [33%], formaldehyde [4%], and NaCl [0.85%]; (ii) formic acid [30%], formaldehyde [4%], and water; and (iii) nitric acid [5%] (3–5) for 24 h ($n = 6$), 48 h ($n = 6$), 72 h ($n = 6$), and 96 hours ($n = 6$). After exposure to the decalcifying solutions, the tissues were routinely processed and again stained with Perls stain for hemosiderin.

The slides were examined blind by one of the authors (RWB), and the intensity of iron staining was scored semiquantitatively as 0 (none), 1+ (mild), 2+ (moderate), or 3+ (marked) based on the examination of 10 random high power fields (HPFs). The results were analyzed using Spearman's rank correlation coefficient.

Results

Both lung and liver tissues exposed to formic acid, formaldehyde, and NaCl decalcifying solution reduced in staining intensity from 2+ to 0.5+ over 96 h (Fig. 1). Suspension in formic acid, formaldehyde, and water reduced lung staining from 3+ to 1+, and liver staining from 2+ to 1+. The intensity of staining of lung tissue immersed in nitric acid reduced from 2+ to 0–0.5+ over the same time, with liver staining decreasing from 3+ to 0. Although there were some instances where staining did not decrease between individual time points, the trend in all samples and in all solutions over the course of the entire experiment (96 h) was for reduction in the intensity of hemosiderin staining. While the reduction in hemosiderin staining failed to reach statistical significance in the specimens decalcified in formic acid, formaldehyde, and NaCl and in formic

[1]Discipline of Pathology, The University of Adelaide, Frome Rd, Adelaide, SA 5005, Australia.
[2]Forensic Science SA, 21 Divett Place, Adelaide, SA 5000, Australia.
Received 28 March 2009; and in revised form 1 July 2009; accepted 3 July 2009.

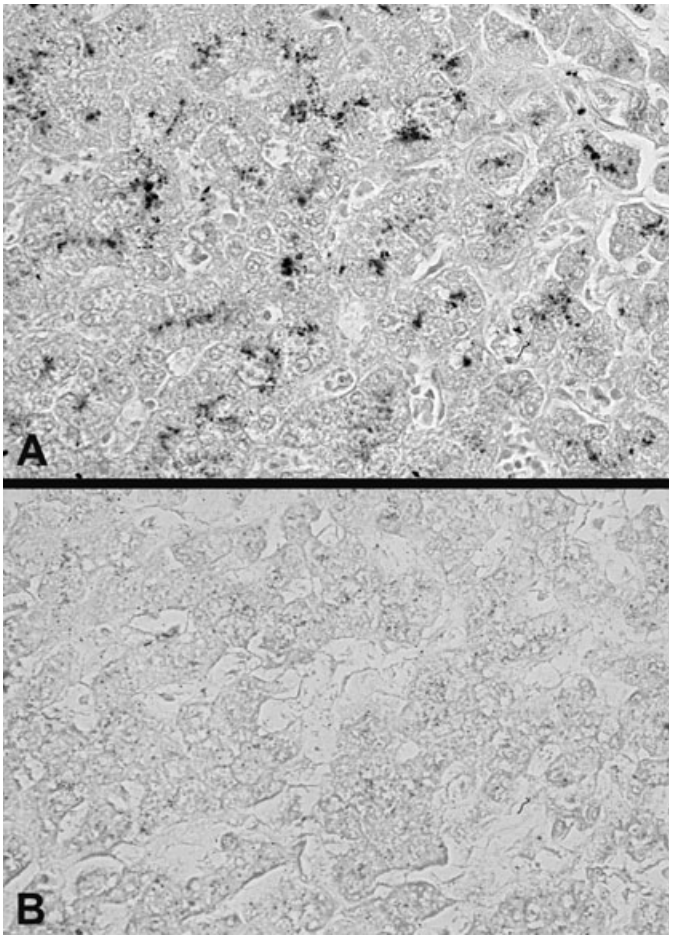


FIG. 1—*(A) A section of control liver demonstrating 3+ staining for hemosiderin (Perls stain ×200). (B) A section from the same block of control liver demonstrating no staining for hemosiderin after exposure to 72 h of standard nitric acid decalcification solution (Perls stain ×200).*

acid, formaldehyde, and water, the trend was significant in tissues decalcified in nitric acid ($p < 0.05$). The results are shown in greater detail in Fig. 2 and in Table 1.

Discussion

Injury to body organs and tissues may have a variety of effects depending on the nature of the trauma. For example, blunt impact to soft tissues may cause tissue damage with extravasation of blood from vessels. This may be observed macroscopically as a contusion or bruise, or histologically as a collection of red blood cells within tissue parenchyma. Trauma to bones may result in breaking or fracture of bony cortices and trabeculae, again with collections of red blood cells in and around the fracture site.

After the escape of blood from vessels following trauma, a series of changes occur that start with migration of polymorphonuclear leukocytes into the area of injury, followed by red cell break down, macrophage migration, fibroblast proliferation, and eventual tissue repair (6). In a medicolegal context, if an injury appears recent, the implication is that only those with recent contact with the victim could have been involved. However, although these stages occur in an orderly manner, the timing of certain stages, such a neutrophil migration, has under certain circumstances been found to be far less precise than textbooks have suggested (2).

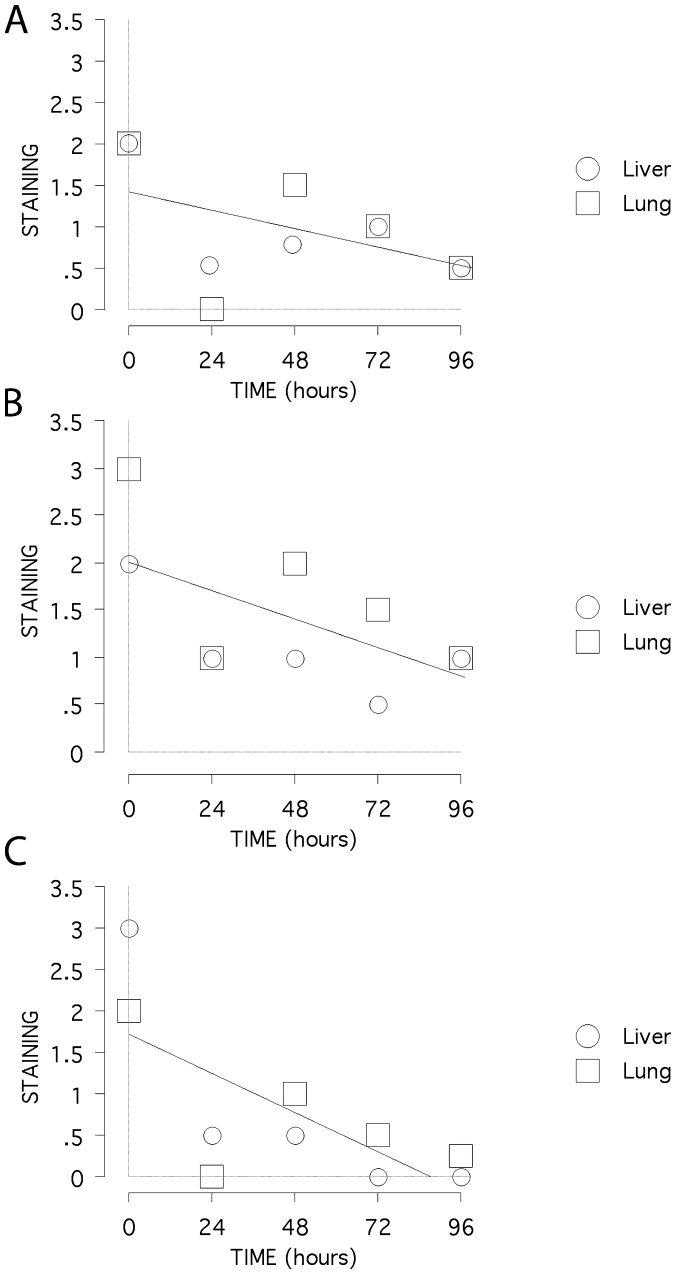


FIG. 2—*Reduction of hemosiderin staining in tissues exposed to three decalcifying solutions over 96 h. (A) formic acid, formaldehyde and NaCl; (B) formic acid, formaldehyde and water; and (C) nitric acid. (staining 0–3+; time in solution—hours).*

TABLE 1—*Semiquantification of hemosiderin staining in lung and liver tissues exposed to three standard decalcifying agents for 24–96 h.*

	Formic Acid, Formaldehyde, & NaCl		Formic Acid, Formaldehyde, & Water		Nitric Acid	
	Lung	Liver	Lung	Liver	Lung	Liver
0 h	2+	2	3	2	2	3
24 h	0	0.5	1	1	0	0.5
48 h	1–2	0.5–1	2	1	1	0.5
72 h	1	1	1–2	0.5	0.5	0
96 h	0.5	0.5	1	1	0–0.5	0

0, no staining; 1+, mild; 2+, moderate; 3+, marked.

Hemosiderin is a granular pigment that forms when hemoglobin from red blood cells breaks down. It is found either within macrophages or lying free within tissues. The estimated time taken for hemosiderin to form has varied in the literature and although it has been claimed that it can deposit after death (7) it is usually considered that it forms during life, commencing between 24 and 48 h after red cell extravasation. For example, in a series of infants with acute pulmonary hemorrhage, hemosiderin-containing macrophages were first identified in bronchial lavage fluid 50 h after the hemorrhage. In the same study, cultured macrophages exposed to sheep red cells demonstrated hemosiderin at 72 h (8). In animal models, hemosiderin has first appeared from 24 to 72 h after exposure to blood (9). The finding of hemosiderin in an injury, therefore, implies that the injury is older than a matter of hours. For this reason, it is standard practice to stain tissue sections for hemosiderin in cases of possible inflicted injury to assist in the assessment of the age of bruises and other injuries.

The pathological evaluation of the age of rib and other fractures in infants and children is often difficult, as osseocartilaginous repairs take a number of days before they are initiated and so histologic changes may not be detected for some time in a recent fracture. For this reason, radiological examination of fractures is an essential component of bony injury assessment in the young. As a component of the histologic assessment of fractures, Perls staining for hemosiderin is also performed looking for iron deposition in tissues in and around fracture sites to determine whether the fracture might be >24 h of age. However, a problem with the histologic examination of material containing bone is the need for decalcification of tissues with acid solutions to enable routine processing and cutting.

There has been some debate in the literature as to the effect of decalcifying agents on iron staining, with some authors finding a reduction in stainable iron following decalcification (10,11), contrasting with others who found iron stores to be better preserved in bone specimens (12). It is also possible that the amount of stainable iron that remains depends on the strength of the decalcifying acid or the type of fixative used (13).

In the current study, we wanted to assess the possible effects of three standard decalcifying agents on hemosiderin stores in lungs and liver. Although the number of cases tested is not large, the study has shown a consistent decrease in hemosiderin staining because of exposure to decalcifying agents; that is, despite some variation in staining color over time, which was possibly affected by sampling, the amount of stainable iron decreased after exposure to all three agents, with the intensity of staining consistently reducing over 96 h from an initial 2–3+ to a final 0–1+ (Table 1). Whether this is caused by an actual reduction in the amount of hemosiderin in the tissues because of a direct action of the decalcifying solutions or whether it instead represents alteration in staining properties is unclear. What the study does show, however, is that the amount of stainable hemosiderin in tissues may be significantly altered by decalcification; the absence of hemosiderin in tissues adjacent to decalcified fracture sites does not, therefore, necessarily indicate that an injury is acute.

References

1. Langlois NEI. The science behind the quest to determine the age of bruises – a review of the English language. Forensic Sci Med Pathol 2007;3:241–51.
2. Byard RW, Wick R, Gilbert JD, Donald T. Histologic dating of bruises in moribund infants and young children. Forensic Sci Med Pathol 2008;4:187–92.
3. Clayden AC. A discussion on the preparation of bone sections by paraffin wax method with special reference to the control of decalcification. J Med Lab Technol 1952;10:103–23.
4. Bancroft JD, Stevens A. Theory and practice of histological techniques, 2nd edn. London: Churchill Livingstone, 1982;305–6.
5. Drury RAB, Wallington EA. Carleton's histological techniques, 4th edn. New York, NY: Oxford University Press, 1973;140–1.
6. Saukko P, Knight B. The pathology of wounds Ch 4. In: Knight's forensic pathology, 3rd edn. London: Arnold, 2004;136–73.
7. Stewart S, Fawcett J, Jacobson W. Interstitial haemosiderin in the lungs of sudden infant death syndrome: a histological hallmark of 'near-miss' episodes? J Pathol 1985;145:53–8.
8. Sherman JM, Winnie G, Thomassen MJ, Abdul-Karim FW, Boat TF. Time course of hemosiderin production and clearance by human pulmonary macrophages. Chest 1984;86:409–11.
9. Epstein CE, Elidemir O, Colasurdo GN, Fan LL. Time course of hemosiderin production by alveolar macrophages in a murine model. Chest 2001;120:2013–20.
10. DePalma L. The effect of decalcification and choice of fixative on histiocytic iron in bone marrow core biopsies. Biotech Histochem 1996;71:57–60.
11. Fong TP, Okafor LA, Thomas W, Westerman MP. Stainable iron in aspirated and needle-biopsy specimens of marrow: a source of error. Am J Hematol 1977;2:47–51.
12. Krause JR, Brubaker D, Kaplan S. Comparison of stainable iron in aspirated and needle-biopsy specimens in bone marrow. Am J Clin Pathol 1979;72:68–70.
13. Stuart-Smith SE, Hughes DA, Bain BJ. Are routine iron stains on bone marrow trephine biopsy specimens necessary? J Clin Pathol 2005;58:269–72.

Additional information and reprint requests:
Roger W. Byard, M.D.
Professor
Discipline of Pathology
Level 3 Medical School North Building
The University of Adelaide
Frome Road
Adelaide 5005
Australia
E-mail: byard.roger@saugov.sa.gov.au

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01409.x
Available online at: interscience.wiley.com

TECHNICAL NOTE

PATHOLOGY AND BIOLOGY

Renato Evando M. Filho,[1,2] *M.S; José Júlio C. Sidrim,*[1,2,3] *Ph.D.; Rossana de A. Cordeiro,*[1,2,3] *Ph.D.; Erica P. Caetano,*[1,2] *B.Sc.; Marcos Fabio G. Rocha,*[1,2,4] *Ph.D.; and Raimunda Sâmia N. Brilhante,*[1,2,3] *Ph.D.*

Trichophyton Mentagrophytes Perforates Hair of Adult Corpses in the Gaseous Period*

ABSTRACT: Despite the substantial literature on mycology, there are still limited reports of the interaction between fungi and human hosts in the postmortem period. Thus, the main goal of this study was to investigate the *in vitro* perforation test using *Trichophyton mentagrophytes* on hair from adult corpses in the postmortem period (gaseous period). The protocol was carried out with positive (prepubescent children's hair) and negative controls (healthy adult hair) as well. One strain of *Trichophyton rubrum* was also used as a negative perforation control. Perforations were found in all the hair samples from corpses and prepubescent children after 12–14 days exposure to *T. mentagrophytes* and were absent in the hair samples of healthy adults. Furthermore, hair perforation was not observed with *T. rubrum*. Our preliminary findings suggest the use of *T. mentagrophytes* as a potential marker of the death interval in forensic science.

KEYWORDS: forensic science, forensic mycology, *Trichophyton mentagrophytes, in vitro* hair perforation, gaseous period, markers of time of death

Advances in forensic medicine and the use of biological markers have improved the postmortem analysis of corpses, which is exemplified by the recent findings versus the estimation of the time of death based only on temperature at different body sites (1,2). Thus, there is a need for wider application of bio- and thanatochemical analyses to estimate precisely the postmortem period. The use of sophisticated models (3) appears to be efficient in estimating time of death, and its combination with biomarkers, e.g., flies (4), may improve forensic techniques. Many other agents may be candidates for estimating the time of death.

The fungus *Trichophyton mentagrophytes* is commonly isolated for laboratory diagnosis, such as in the *in vitro* hair perforation test (5). This test, commonly performed in many mycology laboratories, uses the head hair of blond prepubescent children as a positive control. In this age group, the hair is generally fine and soft, with no medulla, which facilitates verification of the capacity of *T. mentagrophytes* to perforate the hair, which is not observed in the hair from healthy adults because of its higher resistance and the influence of hormonal and immunological factors.

Based on the gradual absence of immunological factors after death, we considered the possibility of using *T. mentagrophytes* to estimate the postmortem period by a hair perforation test. Thus, the main goal of this study was to investigate the *in vitro* perforation test using *T. mentagrophytes* on hair from adult corpses in the postmortem period (gaseous period).

[1]Specialized Medical Mycology Center, Federal University of Ceará, Fortaleza, CE, Brazil.
[2]Postgraduate Program in Medical Microbiology, Federal University of Ceará, Fortaleza, CE Brazil.
[3]Postgraduate Program in Medical Sciences, Federal University of Ceará, Fortaleza, CE Brazil.
[4]Postgraduate Program in Veterinary Science, State University of Ceará, Fortaleza, CE Brazil.

*Supported by CNPq—Conselho Nacional de Desenvolvimento Científico e Tecnológico and by FUNCAP—Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico Process: 9053/08.

Received 23 Sept. 2008; and in revised form 29 June 2009; accepted 2 July 2009.

Materials and Methods

Ethical Aspects

The present study was approved by the Research Ethics Committee of the State University of Ceará (approval under number 064969333-9).

Hair Samples

Scalp hair samples were collected from three different groups for the hair perforation test as follows: (i) hair from corpses (n = 12; victims ranging in age from 18 to 35 years old), (ii) hair from healthy adults (n = 12; with ages ranging from 18 to 35 years old; negative control), and (iii) hair from blond prepubescent children (n = 4; positive control). The hair samples were collected with forceps and a sterile scalpel. None of the living subjects reported recent use of antifungal medicines or similar scalp treatments. The corpses were all from violent death victims, without report of clinical infection caused by fungal microorganisms, and were selected in the gaseous period (between 3 and 5 days after death).

Selections of Strains of T. mentagrophytes *and* T. rubrum

Four strains of *T. mentagrophytes* and one strain of *T. rubrum* (used as a negative control), belonging to the collection of the Specialized Medical Mycology Center (Department of Pathology and

Forensic Medicine of Federal University of Ceará, Brazil), were used in this study.

Exposure to T. mentagrophytes *and to* T. rubrum

Each strain was grown in a small Petri dish (7 cm diameter × 15 mm height) containing unamended agar (Bacto™ Agar; Becton Dickinson Microbiology Systems, Cockeysville, MD). Hair samples were placed on the agar surface with the appropriate fungal strain, according to the previously described protocol for each strain (6). The dishes were then incubated at 25–28°C for up to 30 days. Dishes were evaluated daily; and for each analysis, a part of the hair sample was removed together with the fungi for staining with lactophenol cotton blue (0.05%). Samples were analyzed by light microscopy at 100× and 400× magnification. The test was considered positive when perforation was observed in at least half the thickness of the hair (6).

Results

There were perforations noted in the hair from all the corpses and positive controls (hair from children) exposed to *T. mentagrophytes.* This positive result occurred between days 12 and 14 in children, and 13 and 14 in corpses, after exposure to *T. mentagrophytes* (Table 1).

The microscopic characteristics of the positive samples were noted as the presence of a perforated area in the marrow and/or cortex of the hair penetrating at least half of the hair structure (Fig. 1). The hair samples from the 12 healthy adults (negative control) did not show positive perforation during the 30 days of observation, in any combination of samples and fungal strains. Further, *T. rubrum* was not associated with hair perforation in any case.

TABLE 1—In vitro *hair perforation by* Trichophyton mentagrophytes.

Hair Origin	Strain	Perforation Day (+)	Microscopic Characteristic
Corpses			
1	CEMM 03-3-018	13th	Perforation of the cortex
2	CEMM 03-3-018	14th	Perforation of the cortex
3	CEMM 03-3-018	13th	Perforation of the cortex
4	CEMM 01-4-194	14th	Perforation of the cortex
5	CEMM 01-4-194	14th	Perforation of the marrow
6	CEMM 01-4-194	13th	Perforation of the cortex
7	CEMM 01-5-038	13th	Perforation of the cortex
8	CEMM 01-5-038	13th	Perforation of the cortex
9	CEMM 01-5-038	13th	Perforation of the cortex
10	CEMM 01-1-015	14th	Perforation of the cortex
11	CEMM 01-1-015	14th	Perforation of the cortex
12	CEMM 01-1-015	14th	Perforation of the cortex
Child			
1	CEMM 03-3-018	12th	Perforation of the cortex
2	CEMM 01-4-194	14th	Perforation of the cortex
3	CEMM 01-5-038	13th	Perforation of the marrow
4	CEMM 01-1-015	14th	Perforation of the cortex
Healthy adults			
1	CEMM 03-3-018	–	–
2	CEMM 03-3-018	–	–
3	CEMM 03-3-018	–	–
4	CEMM 01-4-194	–	–
5	CEMM 01-4-194	–	–
6	CEMM 01-4-194	–	–
7	CEMM 01-5-038	–	–
8	CEMM 01-5-038	–	–
9	CEMM 01-5-038	–	–
10	CEMM 01-1-015	–	–
11	CEMM 01-1-015	–	–
12	CEMM 01-1-015	–	–

Discussion

The *in vitro* hair perforation test, initially described by George and Ajello (7), is based on the ability of *T. mentagrophytes* to perforate hair. The standard test is effective and simple to perform, and it is commonly used on blond children's hair because of their low natural resistance and the absence of saturated fatty acid chains, which allow perforation.

Although the defense barriers of humans cease to function after death, this does not happen instantaneously (8). It occurs as part of a natural process of decomposition that involves cell death and the proliferation of microorganisms during phases after death, which are generally divided into four periods: greenish discoloration period (from 24 to 36 h), gaseous period (in days), deterioration period (in months), and skeletonization period (in years) (9). Once started, the various phenomena of cadaver degeneration occur continuously. Besides the hormonal influence during life, the morphology of the hair also undergoes changes after death (10). The cuticle, which during life works as a protective barrier, degenerates after death, allowing attack by microorganisms such as bacteria and fungi.

Despite the substantial literature on mycology, there are still limited reports of the interaction between fungi and human hosts in the postmortem period (11). We have observed the action of hair-perforating enzymes of *T. mentagrophytes* on the scalp hair of adult corpses in comparison with the hair from the healthy adult. The perforation occurs between the 13th and 14th day of incubation, i.e., the gaseous period of adult cadaver decomposition. This indicates that the decay of immunological factors occurs during the gaseous postmortem period, besides the natural after-death degradation of the hair.

In summary, our results open perspectives for an effective characterization of the decay of barriers in the fungus–host interaction after death. Further studies determining the potential of fungal

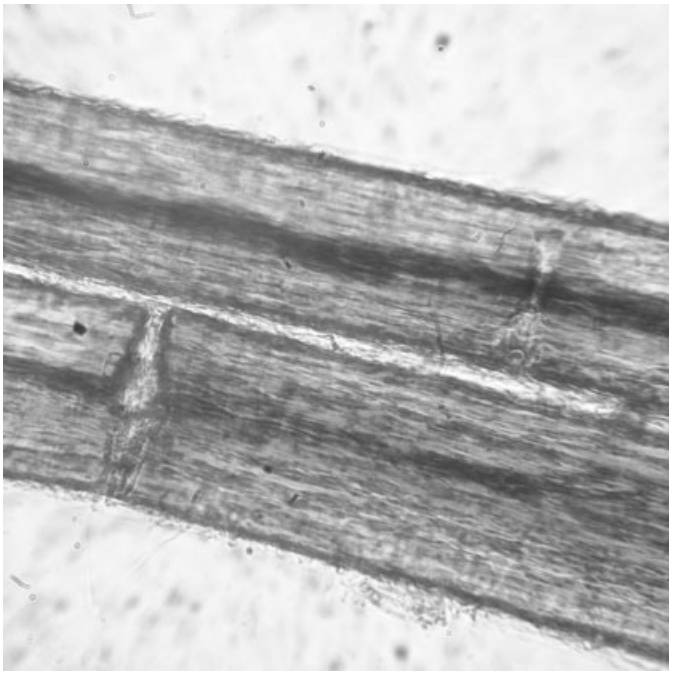


FIG. 1—*The microscopic characteristics of a positive sample with the presence of a perforated area in hair.*

species like *T. mentagrophytes* for use as a marker of the time of death could be interesting for forensic science.

Acknowledgments

This work was supported by the National Research Council (CNPq) and by the Ceará State Foundation for Development of Science and Technology (FUNCAP), Process 9053/08.

References

1. Kaliszan M, Hauser R. Estimation of the time of death based on the measurements of the eye temperature in comparison with other body sites. Arch Med Sadowej Kryminol 2007;57:399–405.
2. Verica P, Janeska B, Gutevska A, Duma A. Post mortem cooling of the body and estimation of time since death. Soud Lek 2007;52:50–6.
3. Mall G, Eisenmenger W. Estimation of time since death by heat-flow Finite-Element model part II: application to non-standard cooling conditions and preliminary results in practical casework. Leg Med 2005;7:69–80.
4. Manlove JD, Disney RH. The use of *Megaselia abdita* (Diptera: Phoridae) in forensic entomology. Forensic Sci Int 2008;175:83–4.
5. Salkin IF, Hollick GE, Hurd NJ, Kemna ME. Evaluation of human hair sources for the *in vitro* hair perforation test. J Clin Microbiol 1985;22: 1048–9.
6. Brilhante RSN, Cordeiro RA, Rocha MFG, Monteiro AJ, Meireles TEF, Sidrim JJC. Tinea capitis in a dermatology center in the city of Fortaleza, Brazil: the role of *Trichophyton tonsurans*. Int J Dermatol 2004;43:575–9.
7. Ajello L, Georg LK. *In vitro* hair cultures for differentiating between atypical isolates of *Trichophyton mentagrophytes* and *Trichophyton rubrum*. Mycopathol 1957;8:3–17.
8. Coe JI, Curran WJ. Definition and time of death. In: Curran WJ, McGarry AL, Petty CS, editors. Modern legal psychiatry and forensic science. Philadelphia, PA: E.A. Davis, 1980;141–64.
9. Di Maio DJ, Di Maio VJ. Forensic pathology, 2nd edn. Boca Raton, FL: CRC Press Inc., 1993.
10. Collier JH. Estimating the post mortem interval in forensic cases through the analysis of post mortem deterioration of human head hair [Master of Arts]. Baton Rouge, LA: Louisiana State University, 2005.
11. Carter DO, Tibbett M. Taphonomic mycota: fungi with forensic potential. J Forensic Sci 2003;48:168–71.

Additional information and reprint requests:
Raimunda Sâmia N. Brilhante, Ph.D.
Rua Barão de Canindé, 210
Montese. CEP: 60.425-540
Fortaleza, CE
Brazil
E-mail: brilhante@ufc.br

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01450.x
Available online at: interscience.wiley.com

TECHNICAL NOTE

PATHOLOGY/BIOLOGY

Hong Huang,[1] Ph.D.; Youyi Yan[1]; Zhong Zuo,[2] Ph.D.; Lin Yang,[1] M.Sc.; Bin Li,[1] Ph.D.; Yu Song,[1] M.Sc.; and Linchuan Liao,[1] Ph.D.

Determination of Adenosine Phosphates in Rat Gastrocnemius at Various Postmortem Intervals Using High Performance Liquid Chromatography

ABSTRACT: Although the change in adenosine phosphate levels in muscles may contribute to the development of rigor mortis, the relationship between their levels and the onset and development of rigor mortis has not been well elucidated. In the current study, levels of the adenosine phosphates including adenosine triphosphate (ATP), adenosine diphosphate (ADP), and adenosine monophosphate (AMP) in gastrocnemius at various postmortem intervals of 180 rats from different death modes were detected by high performance liquid chromatography. The results showed that the levels of ATP and ADP significantly decreased along with the postmortem period of rats from different death mode whereas the AMP level remained the same. In addition, it was found that changes in the ATP levels in muscles after death correlated well with the development of rigor mortis. Therefore, the ATP level could serve as a reference parameter for the deduction of rigor mortis in forensic science.

KEYWORDS: forensic science, rigor mortis, adenosine triphosphate, adenosine diphosphate, adenosine monophosphate, high performance liquid chromatography

Rigor mortis is an important early postmortem change that occurs soon after death and is often used to deduce time and mode of death in forensic science (1). It occurs when muscles of a human are stiff and gradually becomes more severe along with the time after death, until the entire large joint is spastic. Many researchers have shown that the onset and development of rigor mortis can be affected by various factors, such as temperature, pH in the muscles, exercise preceding death, and the cause of death (2,3). There are several mechanisms suggested for the rigor mortis development. Adenosine triphosphates, such as ATP, have been suggested to be the basic biochemical needed for the formation of rigor mortis, which is the same as that of muscle contraction. It is believed that ATP is continuously produced and consumed until death, which will lead to the dramatic decrease of ATP level in the body. When the level of ATP reduces to a certain extent, the actin and myosin filaments combine irreversibly to make muscle maintain rigor solidification. At this stage, rigor mortis sets in. So far, only limited studies reported the relationship between rigor mortis and ATP level. One study demonstrated that the levels of adenosine triphosphate differ a lot between antemortem and postmortem bloodstains prepared from the blood of volunteers and corpses (4). Another study shows that postmortem changes of adenosine nucleotides levels vary significantly in different muscles (5). However, the relationship between ATP level and the onset and development of rigor mortis has not been well elucidated. The purpose of this study was to determine whether the changes of adenosine nucleotide levels in postmortem muscle at the different time periods correlate with the onset and development of rigor mortis by detecting ATP, adenosine diphosphate (ADP), and adenosine monophosphate (AMP) level changes with a high performance liquid chromatography (HPLC) method with modifications (6,7).

[1]West China Preclinical Medical and Forensic Medical School of Sichuan University, Chengdu, Sichuan 610014, China.
[2]School of Pharmacy, Faculty of Medicine, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong SAR.

Received 30 Mar. 2008; and in revised form 5 Aug. 2009; accepted 9 Aug. 2009.

Materials and Methods

Animal Treatments and Muscle Samplings

One hundred and eighty healthy Wistar rats ranging in weight from 230 to 280 g (12–13 weeks) were provided by Huaxi Animal Experiment Center Sichuan University. They were assorted into three groups randomly (gender was not restricted) and killed in different ways. The first group of rats received neck broken (BG). The second group of rats were injected intraperitoneally with a total of 20–30 mg/kg 2% barbitone followed by suffocating via separating and ligating the trachea (SG). The third group of rats had the femoral artery cut off (CG). The corpses of the rats were placed at an average temperature between 12 and 14°C. A piece of gastrocnemius was removed from the rats by blade at 0, 0.5, 1, 1.5, 2, 3, 4, 6, 8, 12, 18, and 24 h after death. The sample was frozen instantly in liquid nitrogen until analysis.

Instrument and Reagent

The HPLC system (Agilent Technologies HP 1100 series) consisted of a quaternary pump, a vacuum degasser, and DAD-detector

(Agilent Technologies Inc., Santa Clara, CA). High-speed and refrigerated centrifuge Eppendorf 5810R (Eppendorf, Hamburg, Germany) was used for the sample preparation. Standards of ATP (5′-triphosadenine natrium salt), ADP (5′-adenosine diphosphate natrium salt), and AMP (5′-adenosine monophosphate natrium salt) were obtained from Sigma Corporation (Santa Clara, CA). Other reagents were analytical pure and made in China. Water used for the experiments was prepared by Ultrapure water demineralizer (Millipore, Billerica, MA).

Chromatographic Condition

Compounds were separated by Shim-pack vp-ODS C_{18} column (Shamadizu Japan, 5 μm, 250 mm × 4.6 mm) eluted with 100% of 50 mM potassium phosphate buffer (pH 6.5) at a flow rate of 1 mL/min. The column temperature was set at 20°C. The UV detection wavelength was 254 nm and the spectrum was acquired by DAD. Sample injection volume is 20 μL.

Sample Preparations

The collected rat gastrocnemius muscle samples were prepared as follows: (i) Sample was taken out from liquid nitrogen, weighed by electronic scale, and placed into precooled glass mortar immediately. (ii) Precooled 0.4 mol/L perchloric acid (5 mL/g) was added in the mortar followed by homogenization in an iced bath. (iii) The homogenate was centrifuged at 4,000 rpm for 10 minutes at 4°C. (iv) The supernatant was collected and mixed with same volume of 1 mol/L potassium dihydrogen phosphate solution followed by adjusting the pH to 6.5. (v) Repeated step (iii). (vi) The collected supernatant was combined and stored in a −20°C freezer. (vii) At the time of analysis, the frozen supernatant was thawed at room temperature and filtered with a 0.45 μm filter unit prior to injection into HPLC for analysis.

Results

Qualitative Analysis

Mixture of ATP, ADP, and AMP standards and prepared samples were analyzed under the above-mentioned HPLC conditions. As shown in Figs. 1 and 2, the retention times of the three adenosine phosphates in the samples were similar to that obtained from the authentic standards. The retention times for ATP, ADP, and AMP were at 9.535, 11.800, and 19.102 min, respectively. The whole HPLC run lasted 24 min for each sample injection. The peaks of three adenosine phosphates were further confirmed by spiking in authentic standard mixtures. As demonstrated in Fig. 3, chromatographic peaks of ATP, ADP, and AMP were elevated when standard mixture was spiked into the same samples. In addition, the ultraviolet spectrograms of sample and standards were very similar. These results suggested that we could detect ATP, ADP, and AMP levels in rat gastrocnemius samples by the current HPLC/DAD method.

Calibration Curves

Under the current HPLC conditions, external reference method was used to quantitate the concentration of adenosine nucleotides in prepared samples. The concentration range of ATP, ADP, and AMP is 2.15∼344 mg/L, 4.5∼360 mg/L, and 1.475∼236 mg/L, respectively. Under these concentration ranges, linear correlation between concentration and peak area is satisfactory. The standard curves of ATP, ADP, and AMP are expressed as follows: ATP: Y = 19274X+44.982 (R = 0.9991, n = 6); ADP: Y = 22085X+203.37 (R = 0.9991, n = 6); AMP: Y = 23772X+184.7 (R = 0.999, n = 6).

Determination of Adenosine Phosphates

After injecting 20 μL of prepared samples into HPLC, the concentrations of ATP and ADP were calculated according to the related standard curves. The result of concentrations from each group is expressed as mean ± SD. Regression analysis and q-test between-groups and within-groups were carried out by using SPSS 12.0. A $p < 0.05$ was considered statistically significant.

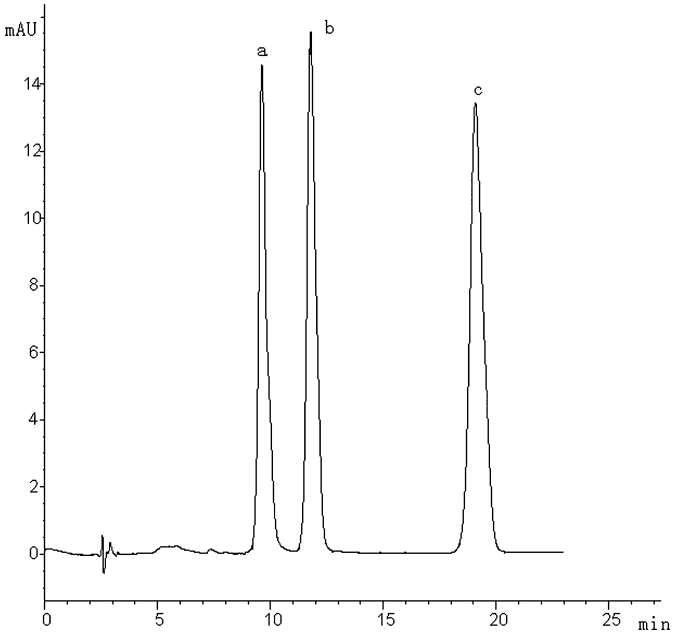


FIG. 1—*HPLC chromatogram of ATP, ADP, and AMP standard mixture (a. ATP, b. ADP, c. AMP).*

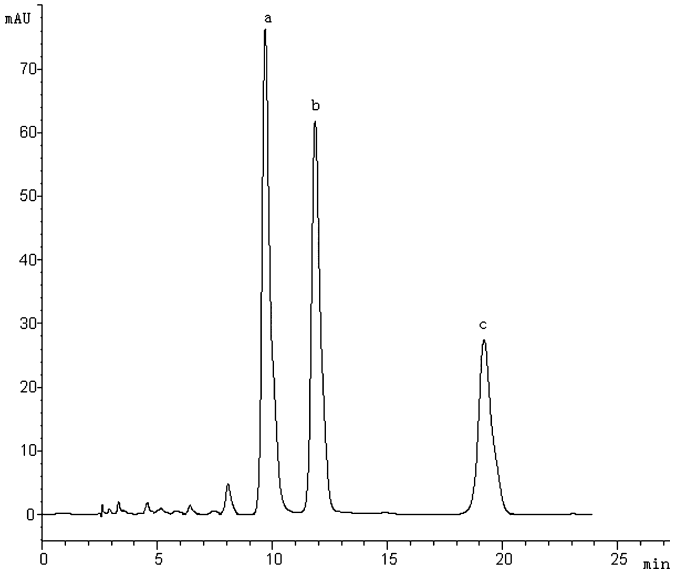


FIG. 2—*HPLC chromatogram of ATP, ADP, and AMP of rat gastrocnemius sample (a. ATP, b. ADP, c. AMP).*

Change of ATP Levels

As shown in Table 1, the levels of ATP in rat gastrocnemius decreased gradually after death. ATP could be detected until 6-h postmortem in both CG and SG group and until 8-h postmortem in BG group. It was noticed that the ATP concentration in rat gastrocnemius was significantly different at each time-point among the three studied groups ($p < 0.05$). Meanwhile, the difference between different time-points is also significant within each treatment group ($p < 0.05$). As shown in Figs. 4, 5, and 6 (CG: ATP, $\hat{y} = 2.983 - 0.591x$, r = 0.901; SG: ATP, $\hat{y} = 2.647 - 0.486x$, r = 0.939, BG: ATP, $\hat{y} = 3.477 - 0.490x$, r = 0.940.), the ATP levels in rat gastrocnemius from all three groups showed a distinct trend of decrease with time.

Change of ADP Levels

As also indicated in Table 1, the ADP contents in rat gastrocnemius from each group differs significantly at different time-points ($p < 0.05$). The ADP contents at 0.5-h, 1.0-h, 12.0-h postmortem is not significant among the three treatment groups, whereas they differed significantly at other time-points postmortem ($p < 0.05$). As shown in Figs. 7, 8, and 9 (CG: ADP, $\hat{y} = 0.733 - 0.035x$, r = 0.702, SG: ADP, $\hat{y} = 0.710 - 0.031x$, r = 0.718, BG: ADP: $\hat{y} = 0.930 - 0.037x$, r = 0.714.), the contents of ADP in rat gastrocnemius from the three groups also demonstrated a distinct trend of decrease with time.

Change of AMP Levels

Unlike ATP and ADP, the contents of AMP in rat gastrocnemius from the three groups of rats showed no correlation with time.

Discussion

Rigor mortis is believed to be because of the disappearance of ATP from muscle (8). ATP is the basic source of energy for muscle contraction. In the absence of ATP, actin and myosin filaments become permanently complexed and rigor mortis sets in (8), which is a consequence of thin myofilament moving toward thick

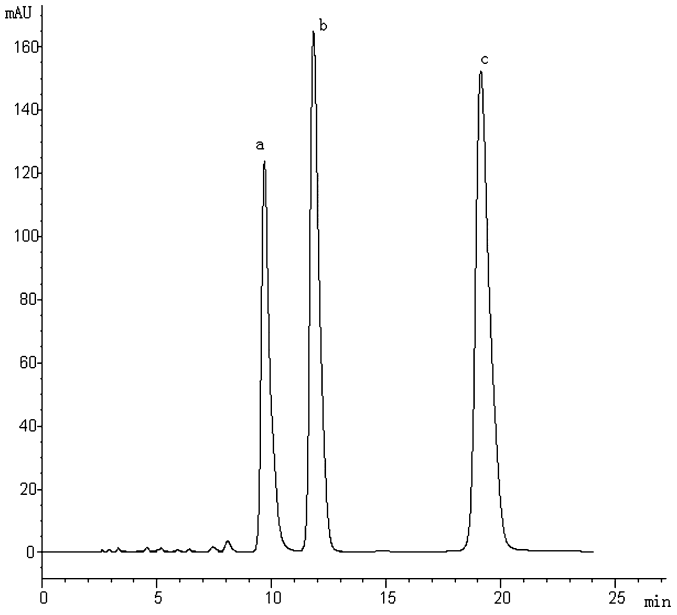


FIG. 3—*HPLC chromatogram of ATP, ADP, and AMP of rat gastrocnemius sample spiked with standard mixture (a. ATP, b. ADP, c. AMP).*

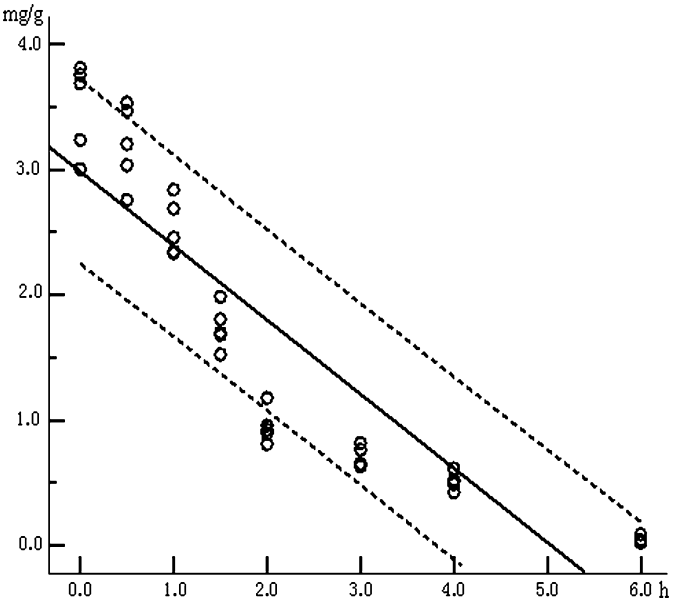


FIG. 4—*Relationship of ATP levels in CG groups with the postmortem time (Solid line: regression line; Dotted line: 90% confidence interval).*

TABLE 1—*Levels of adenine nucleotide orthophosphoric acid at different postmortem time-point in three groups.*

	CG		SG		BG	
Time	ATP (mg/g)	ADP (mg/g)	ATP (mg/g)	ADP (mg/g)	ATP (mg/g)	ADP (mg/g)
0.0 h	3.4982 ± 0.3576	0.8606 ± 0.1663	2.6854 ± 0.0939	0.6351 ± 0.0233	4.1671 ± 0.1193	0.6536 ± 0.0092
0.5 h	3.1389 ± 0.0638	0.8995 ± 0.1186	2.4353 ± 0.0311	0.9522 ± 0.4781	3.3699 ± 0.2246	0.8462 ± 0.0069
1.0 h	2.4721 ± 0.1387	1.1206 ± 0.0474	2.4033 ± 0.1683	0.8533 ± 0.1127	2.8667 ± 0.0256	1.0084 ± 0.0412
1.5 h	1.7391 ± 0.0548	0.9831 ± 0.0660	2.1303 ± 0.2314	0.7132 ± 0.0678	2.0335 ± 0.3672	1.2066 ± 0.0692
2.0 h	0.6367 ± 0.1751	0.5120 ± 0.1696	1.4317 ± 0.2381	0.9717 ± 0.0882	2.6462 ± 0.0515	1.1170 ± 0.0744
3.0 h	0.6143 ± 0.2906	0.4744 ± 0.1053	1.0446 ± 0.3309	0.5993 ± 0.0336	1.8810 ± 0.0335	1.0687 ± 0.0787
4.0 h	0.5072 ± 0.0135	0.4397 ± 0.0370	0.2086 ± 0.0803	0.2602 ± 0.0833	1.3700 ± 0.0222	0.8934 ± 0.0363
6.0 h	0.0453 ± 0.0263	0.1080 ± 0.0514	0.1114 ± 0.0583	0.3328 ± 0.0193	0.2050 ± 0.0791	0.3884 ± 0.0499
8.0 h	ND	0.1429 ± 0.0397	ND	0.1795 ± 0.0203	0.0050 ± 0.0077	0.1955 ± 0.0108
12.0 h	ND	0.1420 ± 0.0643	ND	0.1496 ± 0.0007	ND	0.1757 ± 0.0045
18.0 h	ND	0.1879 ± 0.0039	ND	0.2078 ± 0.0107	ND	0.4509 ± 0.1751
24.0 h	ND	0.1518 ± 0.2628	ND	0.1888 ± 0.0416	ND	0.1754 ± 0.0071

ND, not detectable.

myofilament. Under certain circumstances, cross-bridge of thick myofilament binds with thin myofilament actin in a reversible manner and moves toward M line. Cross-bridge can break down ATP so as to obtain energy to make it move (9). In addition, decomposition of ATP can be affected by temperature and enzymatic activity. At low temperatures, speed of ATP decomposition is very slow (4). Likewise, the speed of ATP decomposition slows down as enzymatic activity decreases (10). The consumption and synthesis of ATP is a dynamic equilibrium process in the living body and such equilibrium no longer exists after death.

The basic biochemical process for the formation of rigor mortis is the same as that of muscle contraction except that the former happens postmortem and the latter antemortem. Liao et al. observed in rat muscle that the striations of muscle were blurred within 4 h and became apparent from 6-h to 24-h postmortem. They also noticed that the distance between two Z lines shortened and the H band narrowed in the process of muscle rigor (11). Using a scanning electron microscope, Wang et al. (12) observed that the length of sarcomere of rigor mortis without disturbance was obviously shorter than that of disturbance and it is the shortest at 5-h postmortem in rat muscles. Kobayashi et al. (2) found in rat muscle that the muscle tension first increased and then decreased over the first 8-h postmortem period, and tension always reached its maximum level earlier than rigidity and decreased more rapidly and markedly than rigidity (13), and the rates of ATP level reduction were different between different muscles (5). These phenomena are

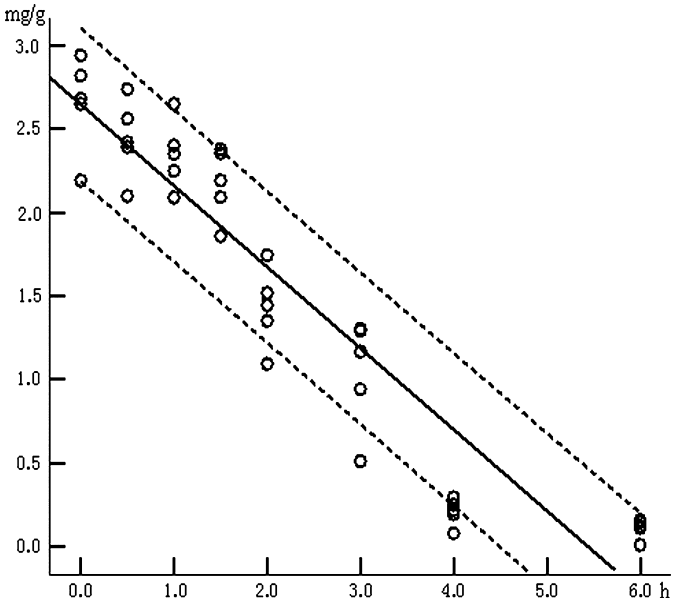


FIG. 5—*Relationship of ATP levels in SG groups with the postmortem time (Solid line: regression line; Dotted line: 90% confidence interval).*

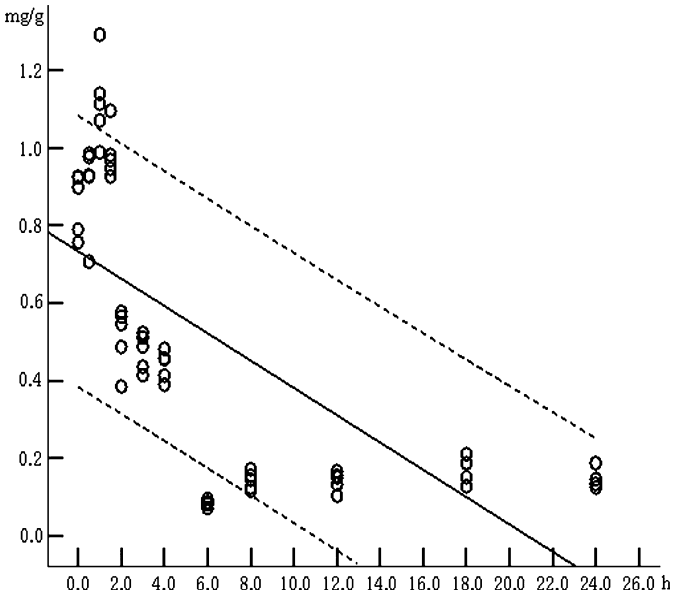


FIG. 7—*Relationship of ADP levels in CG groups with the postmortem time (Solid line: regression line; Dotted line: 90% confidence interval).*

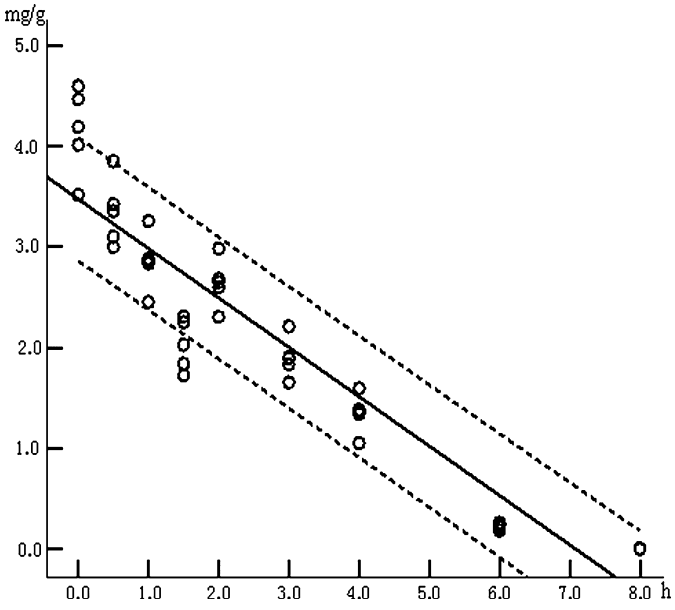


FIG. 6—*Relationship of ATP levels in BG groups with the postmortem time (Solid line: regression line; Dotted line: 90% confidence interval).*

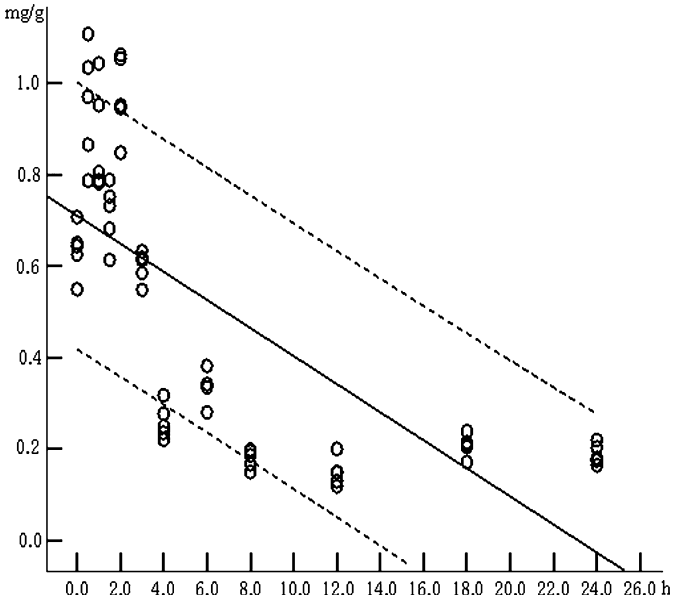


FIG. 8—*Relationship of ADP levels in SG groups with the postmortem time (Solid line: regression line; Dotted line: 90% confidence interval).*

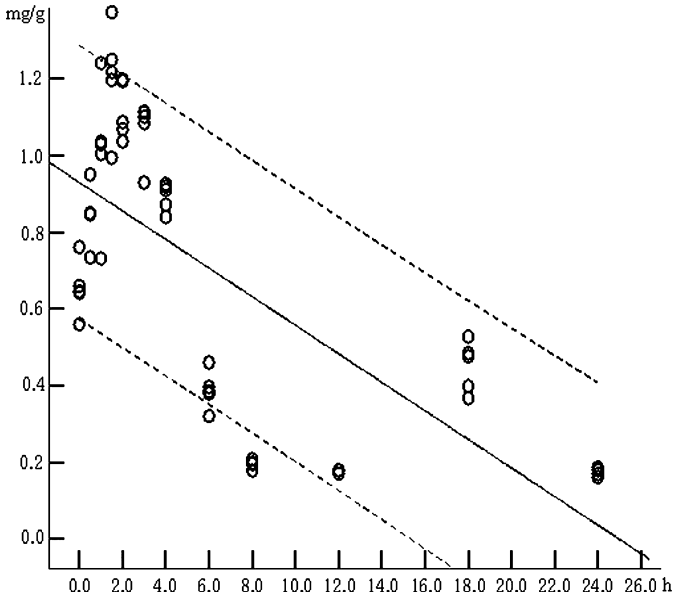


FIG. 9—*Relationship of ADP levels in BG groups with the postmortem time (Solid line: regression line; Dotted line: 90% confidence interval).*

consistent with what we have observed in our experiment. We found that the levels of ATP in the rat gastrocnemius from all treatment groups by different death mechanisms decreased along with time and completely disappeared at 8-h postmortem. When ATP was exhausted under postmortem status, cross-bridge moving stopped and then muscle tension decreased. Therefore, the decrease of ATP level commendably is expected to indicate the occurrence and development of rigor mortis. In addition, ATP may also play a key role in rigidity re-establishment. Krompecher et al. (14) found in rat muscle specimen that a significant rigidity can reappear if the breaking occurs before the process is complete. Therefore, it is suggested that the development of rigidity muscle fibers enter the rigid state progressively according to their different metabolism.

In our experiment, we also found that the ATP level in the gastrocnemius is significantly different among the three treatment groups of different death mechanisms. It was expected that the ATP levels in the gastrocnemius were affected by the causes of death. Rats in the tachygenesis death group such as BG group were terminated in a very short period without any vigorous exercise, and the ATP levels in muscles remained at a relatively high level. Rats in SG group took a much longer period to die and may have experienced vigorous struggle before death, which may consume large amounts of ATP during the process. Therefore, the ATP levels in SG group were much less than that in BG group. For rats in CG group, the duration of death is similar to that in SG group but much longer than that in BG group. However, the rats in CG group would not go through the vigorous struggle as the SG group did. Thus, the ATP levels remaining in the CG group would be higher than that in SG group but less than that in BG group. No matter how, the trend of ATP postmortem level changes are the same for the three treatment groups even though the starting levels of ATP in each group may differ because of different cause of death.

Finally, it was shown in our experiment that the trend of decrease in ADP level with time is similar to that of ATP level, but the former can be detected in all postmortem intervals, whereas the level of ATP can only be detectable until 8-h postmortem, and the level of AMP is not correlated with postmortem time. It is suggested that the changes of ADP and AMP levels do not seem to correlate well with the process of rigor mortis. Therefore, the ATP level is the most relevant parameter to deduce the rigor mortis.

Conclusion

Based on our findings, it is concluded that the levels of ATP were significantly different at different postmortem intervals and demonstrated a distinct trend of decrease with postmortem time. Our results suggest that the change of ATP level correlates well with the development of rigor mortis.

Acknowledgment

The authors would like to thank Miss Ju Jing, Department of Forensic Pathology School of Preclinical Medicine and Forensic Medicine Sichuan University, for helping us to perform the experiments.

References

1. Vain A, Kauppila R, Vuori E. Estimation of the breaking of rigor mortis by myotonometry. Forensic Sci Int 1996;79:155–61.
2. Kobayashi M, Ikegaya H, Takase I, Hatanaka K, Sakurada K, Iwase H. Development of rigor mortis is not affected by muscle volume. Forensic Sci Int 2001;117:213–9.
3. Varetto L, Curto O. Long persistence of rigor mortis at constant low temperature. Forensic Sci Int 2005;147:31–4.
4. Sugie H, Nishikawa T, Funao T. Quantitation of nucleotides, nucleosides and bases in antemortem and postmortem bloodstains by high-performance liquid chromatography. Forensic Sci Int 1995;71:123–30.
5. Kobayashi M, Takatori T, Iwadate K, Nakajima M. Reconsideration of the sequence of rigor mortis through postmortem changes in adenosine nucleotides and lactic acid in different rat muscles. Forensic Sci Int 1996;82:243–53.
6. Wu JH, Tang LF, Tan BY, Ouyang B, Li ZW. Determination of ATP, ADP and AMP in rat myocardium and skeletal muscles by RP—HPLC. J Instrumental Anal 1999;18:55–7.
7. Miao Y, Wang CL, Yin HJ, Shi DZ, Chen KJ. High performance liquid chromatogram (HPLC) determination of adenosine phosphates in rat myocardium. J Peking University (Health Sci) 2005;37:201–2.
8. DiMaio DJ, DiMaio VJM. Forensic pathology, 2nd edn. New York, NY: CRC, 2001.
9. Lõrinczy D, Könczöl F, Farkas L, Belagyi J, Schick C. Nucleotide-induced changes in muscle fibres studied by DSC TMDSC. Thermochimica Acta 2001;377:205–10.
10. Sun DH, Liao ZG, Li QY. Changes of enzyme histochemistry in rat postmortem muscles. J Law Med Sci 1999;6:102–4.
11. Liao ZG, Yi XF, Zhang YL, Peng XM, Li QY. Observations on rat's muscle at various postmortem intervals by scanning electron microscopy. J West China University Med Sci 1998;29:323–5.
12. Wang X, Li M, Liao ZG, Yi XF, Peng XM. Experimental study of restiffening of the rigor mortis. J Forensic Med 2001;17:202–4.
13. Kobayashi M, Takemorib S, Yamaguchi M. Differential rigor development in red and white muscle revealed by simultaneous measurement of tension and stiffness. Forensic Sci Int 2004;140:79–84.
14. Krompecher T, Gilles A, Brandt-Casadevall C, Mangin P. Experimental evaluation of rigor mortis IX. The influence of the breaking (mechanical solution) on the development of rigor mortis. Forensic Sci Int 2008;176:157–62.

Additional information and reprint requests:
Linchuan Liao, Ph.D., Prof.
West China Preclinical Medical and Forensic Medical School
Sichuan University
Chengdu, Sichuan 610014
China
E-mail: linchuanliao@scu.edu.cn

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01431.x
Available online at: interscience.wiley.com

TECHNICAL NOTE

PATHOLOGY AND BIOLOGY; CRIMINALISTICS

Humbert de Freminville,[1] *M.D., M.Sc.; Nicolas Prat,*[1] *M.D., M.Sc.; Frederic Rongieras,*[1] *M.D., M.Sc.; and Eric J. Voiglio,*[1] *M.D., Ph.D., F.A.C.S., F.R.C.S.*

Less-Lethal Hybrid Ammunition Wounds: A Forensic Assessment Introducing Bullet-Skin-Bone Entity

ABSTRACT: Agencies all around the world now use less-lethal weapons with homogeneous missiles such as bean bag or rubber bullets. Contusions and sometimes significant morbidity have been reported. This study focuses on wounds caused by hybrid ammunition with the pathologists' flap-by-flap procedure. Twenty-four postmortem human subjects were used, and lesions caused on frontal, temporal, sternal, and left tibial regions by a 40-mm hybrid ammunition (33 g weight) were evaluated on various distance range. The 50% risk of fractures occurred at 79.2 m/sec on the forehead, 72.9 m/sec on the temporal, 72.5 m/sec on the sternum, and 76.7 m/sec on the tibia. Skin lesions were not predictors of bone fracture. There was no correlation between soft and bone tissue observed lesions and impact velocity (correlated to distance range). Lesions observed with hybrid ammunition were the result of bullet-skin-bone entity as the interaction of the projectile on skin and bone tissues.

KEYWORDS: forensic science, less-lethal weapon, hybrid ammunition, forensic assessment, wounds, bullet-skin-bone entity

Less-lethal weapons and less-lethal projectiles (LLP) have now been commonly used by law enforcement agencies in Europe and all around the world for more than 30 years (1).

Some LLP can only be shot by specially designed launchers such as MR-35 (2) and Flash-Ball® (3), but others are shot by firearms initially designed for lethal ammunitions (4). Well known are square and round beanbags fired by 12G shotgun (5). Various injuries, mainly lung contusions (5), liver contusion (6), and even death (7), have been reported with these LLP. More recently, rubber bullets launched from 40-mm rifle have been introduced to be used by law enforcement agencies (5). These large bore LLP are designed to incapacitate individuals and are supposed not to cause lethal injuries or permanent sequels.

Importance of lesions depending on distance, velocity, or kinetic energy has been well investigated for 12 Gauge homogeneous rubber bullets (8,9).

To have a stable flight, most of 40-mm LLP are hybrid, composed of a rear hard plastic part that takes the spin from the rifling of the barrel and of a soft front part designed to attenuate the impact on the target.

The aim of this preliminary study was to assess whether there was a correlation between velocity at impact of a hybrid 40-mm LLP and severity of lesions observed on postmortem human subjects (PMHS) and whether there was a correlation between skin lesions and underlying bone fractures.

[1]UMRESTTE UMR T 9405, Université de Lyon, Université Claude Bernard Lyon 1, Faculté de Médecine Lyon-Est Claude Bernard, 8, Rockefeller Ave, F-69373 Lyon Cedex 08, France.

Received 3 Mar. 2009; and in revised form 1 June 2009; accepted 12 July 2009.

Materials and Methods

Gun and Ammunitions

A H&K 69 rifled-bore Handgun Launcher (Heckler & Koch, Oberndorf, Germany) was used (Fig. 1). With a break top action and ladder rear sight, the HK69 has a retractable shoulder stock and external hammer. Technical features are : 40 × 46 mm caliber, capacity : 1, mode of fire : manual, width (in.) 2.25, height (in.) 8.00, weight (lb.) 5.77 bbl, length (in.) 14.00, overall length (in.) 28.90.

An experimental less-lethal hybrid ammunition (classified) (Fig. 2) was used for the experiments. The ammunition was a 40-mm cartridge loaded with black powder. The projectile (weight 33 g, length 65 mm) was composed of a foam front part (weight 3 g, length 3.7 mm) and a plastic rear part (weight 30 g, length 28 mm) (Fig. 2).

FIG. 1—*H&K 69 smooth-bore Handgun Launcher.*

Missile Velocity Measurements

Velocity of the hybrid ammunition was measured with a 24 GHz Doppler Radar settled upside the muzzle of Launcher. The start recording came from a microphone ahead Muzzle, and the data were acquired with a Weibel Bay (Weibel Scientific Inc. Alleroed, Denmark).

A ballistic chronograph composed of two photoelectric screens gave the velocity at impact and the start of high-speed video. All signals were recorded at 500 kHz on Numerical NICOLET ODYSSEY Recorder (Nicolet Instrument Technologies, Madison, WI).

A high-speed video Fastcam-APX 120K (Photron USA Inc., San Diego, CA) 10,000 frames/sec was used to evaluate incidence and action of projectile at impact.

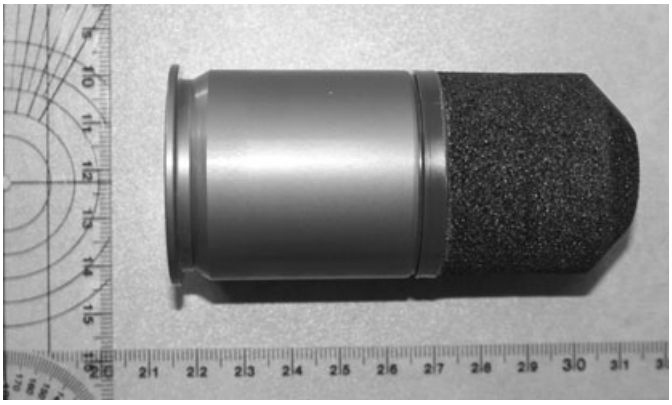


FIG. 2—*Hybrid ammunition 40 × 46 mm.*

PostMortem Human Subjects

Twenty-three unembalmed human corpses were obtained from the anatomic donation program of Lyon (France). All experiments were conducted at the Laboratory of Anatomy of the Faculty of Medicine of Lyons (France) in strict accordance with French laws and regulations. Corpses came from donors who had willingly donated them during their lifetime by testament to science for teaching and research purposes. Ethical practices were followed throughout the testing, with all specimens being treated with respect (5).

Eleven men and 12 women, 74–97 years old range, were placed on a special seat and shot once in the frontal, temporal, sternal, and the tibial areas. A total of 92 lesions were produced (Table 1).

Exclusion Criteria

In case of oblique impact of projectile with lesions mainly caused by the plastic rear part, the lesion was excluded from the study.

Dissections

The superficial skin wound (or the impact points) was used to center the quadrilateral dissection. The edge was drawn at 8 cm from the center, and the skin and subcutaneous fat flap was created and raised as a book page. Underlying bone was examined. Skin lesions were quoted as no wound (uninjured or simple skin abrasion) or wound (linear, curved, or stellar), and bone lesions were quoted as no fracture or fractured (linear, or comminuted).

Statistic Analysis

A logistic regression analysis was performed to establish the probability of bone fracture according to projectile velocity.

TABLE 1—*Velocity of the LLP and situation of the lesions produced.*

					Velocity (m/sec)			
					Target Area			
PMHS #	Sex	Age (years)	Corpulence	Head Perimeter (cm)	Frontal	Temporal	Sternal	Tibial
1	M	82	Thin	55	64.0	67.4	64.3	64.1
2	F	95	Thin	54	64.5	60.6	65.0	72.5
3	F	96	Medium	53	64.4	71.2	65.2	64.6
4	M	63	Thin	57	55.5	51.9	60.0	54.2
5	M	85	Thin	52	62.7	61.9	55.7	57.3
6	M	84	Heavy	58	56.4	58.8	56.2	50.8
7	F	74	Medium	54	49.3	55.3	47.9	53.0
8	M	77	Medium	60	43.1	53.1	48.0	48.2
9	M	83	Heavy	53	53.2	57.7	49.2	52.0
10	M	87	Thin	57	50.8	53.7	92.2	53.5
12	F	89	Medium	54	66.2	77.3	81.0	73.1
13	M	82	Thin	55	80.9	78.6	80.6	74.6
14	M	76	Heavy	56	75.1	80.6	75.7	75.5
15	F	93	Thin	54	62.6	68.1	61.7	78.8
16	F	93	Medium	53	60.1	55.9	52.7	59.0
17	F	96	Thin	50	61.8	53.4	58.4	58.1
18	F	79	Thin	52	69.0	76.2	71.6	76.9
19	F	89	Medium	50	37.9	50.0	47.3	41.5
20	F	97	Thin	53	74.4	75.4	75.1	79.2
21	F	94	Heavy	56	73.9	76.1	69.9	83.9
22	M	81	Heavy	58	83.1	82.4	80.7	79.4
23	M	81	Thin	52.5	82.2	87.5	82.8	91.4
24	F	90	Thin	57	77.4	77.0	80.2	77.3

LLP, less-lethal projectiles; PMHS, postmortem human subjects.

Correlation between skin wound and bone fracture was studied calculating sensitivity and specificity.

Results

Of 92 lesions, two met exclusion criteria (MB7 frontal & MB19 frontal).

Frontal Lesions

Frontal lesions are summarized in Table 2.

Sensitivity of skin wound as predictor of bone fracture was 0.67 and specificity was 0.06.

Probability of frontal bone fracture according to projectile velocity is given in Fig. 3.

Temporal Lesions

Temporal lesions are summarized in Table 3.

Sensitivity of skin wound as predictor of bone fracture was 0.375 and specificity was 0.6.

Probability of temporal bone fracture according to projectile velocity is given in Fig. 4.

Sternal Lesions

Sternal lesions are summarized in Table 4.

Sensitivity of skin wound as predictor of bone fracture was 0.11 and specificity was 1.

Probability of sternal bone fracture according to projectile velocity is given in Fig. 5.

Tibial Lesions

Tibial lesions are summarized in Table 5.

Sensitivity of skin wound as predictor of bone fracture was 0.857 and specificity was 0.688.

TABLE 2—*Skin lesions and bone lesions in frontal region.*

	Wound	No Wound	Total
Fracture	2	1	3
No fracture	17	1	18
Total	19	2	21

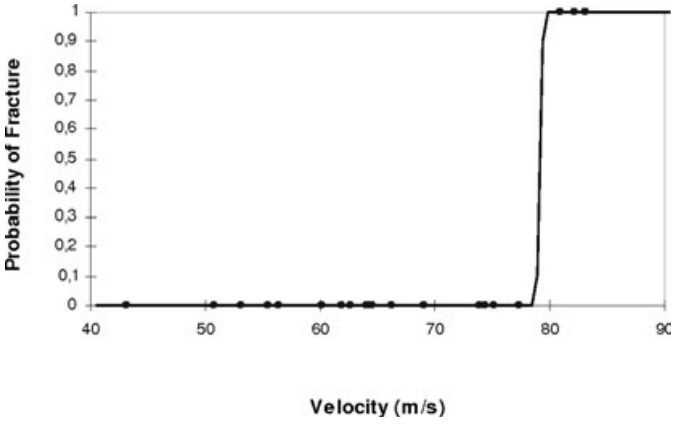


FIG. 3—*Probability of frontal bone fracture according to missile velocity at impact (logistic regression).*

Probability of tibial bone fracture according to projectile velocity is given in Fig. 6.

Discussion

Basic Statement

The 92 impacts recorded allowed the study of four different anatomical site wounds. In these sites, skin lesions were related to subcutaneous and bone lesions. After each serial test—including frontal bone, temporal bone, sternal, and tibial area—the flap-by-flap dissection was performed. These dissections let note differences between visual description, palpation, and the existence of a fracture.

Probability of Fracture

Probability of bone fracture according to projectile velocity was modeled by logistic regression analysis of data, and 50% risk of fractures was defind. The 50% risk of fracture was 79.2 m/sec on

TABLE 3—*Skin lesions and bone lesions in temporal region.*

	Wound	No Wound	Total
Fracture	3	5	8
No fracture	6	9	15
Total	9	14	23

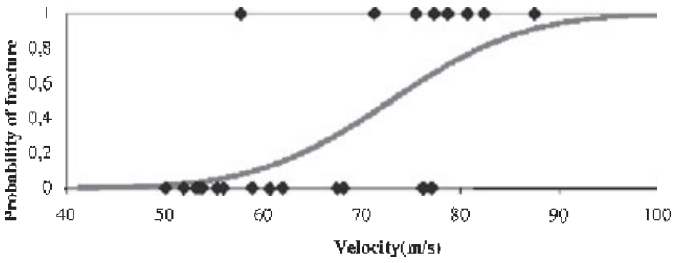


FIG. 4—*Probability of temporal bone fracture according to missile velocity at impact (logistic regression).*

TABLE 4—*Skin lesions and bone lesions in sternal region.*

	Wound	No Wound	Total
Fracture	1	8	9
No fracture	0	14	14
Total	1	22	23

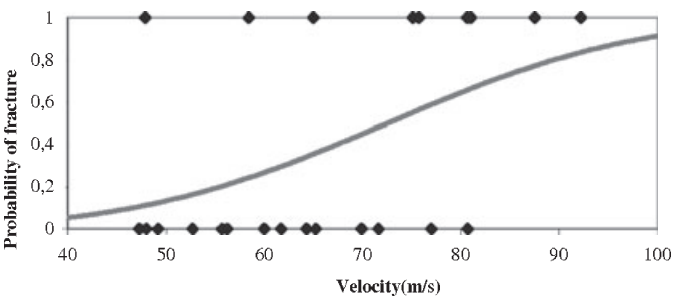


FIG. 5—*Probability of sternal bone fracture according to missile velocity at impact (logistic regression).*

TABLE 5—*Skin lesions and bone lesions in tibial region.*

	Wound	No Wound	Total
Fracture	6	1	7
No fracture	5	11	16
Total	11	12	23

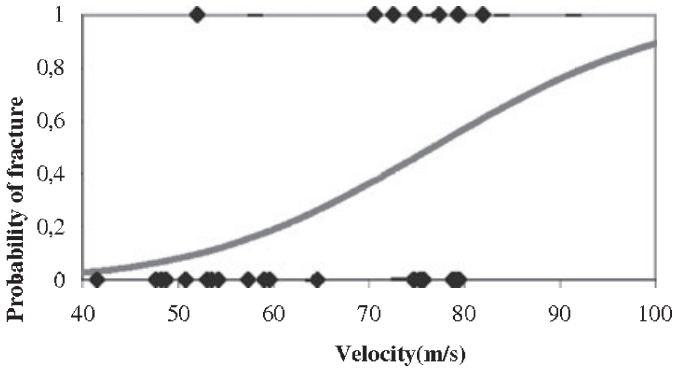


FIG. 6—*Probability of tibial bone fracture according to missile velocity at impact (logistic regression).*

frontal bone (Fig. 3), 72.9 m/sec on temporal bone (Fig. 4), 72.5 m/sec on sternum (Fig. 5), and 76.7 m/sec on tibia (Fig. 6). For all these sites, 50% risks of fracture were in closed values. Except for frontal bone with an apparent threshold for fracture, for other sites, it was impossible to correlate fracture and missile velocity.

Aspect of Skin and Subcutaneous Wounds

Various aspects of skin wound could be observed: abrasion, linear wound, arc-like, and star-like wound. The arc-like aspect was only observed on the frontal area, and this can be explained by the thickness of the skin and missile printing on bone. We noticed with a very low deep impact (37.9 m/sec), an ¾ arc-like wound, and a complete arc-like wound with more than 80 m/sec. In 92 tests, only four MB were woundless, so that means that there are skin wounds in more than 95% of LLP impacts.

We observed abrasions with minor wounds for the majority of the planned area like temporal and sternal sites. Linear wounds were observed on tibial impact: this can be explained easily by the anatomy of the bone, skin is cut on the anterior crest of the shin.

Aspect of Fracture

For 63 impacts (68.4%), there were no fractures. Fractures occured in 33% as star-like for the frontal bone and arc-like for 100% on the temporal bone, especially when the impact was between 78 and 80 m/sec, and higher. Aspect of the fracture was related to solidity of bone: weak for temporal bone and resistant for frontal bone. Tibial fractures were linear (transversal and simple fracture) from an impact of 70–72 m/sec, and comminutive when energy was higher (from 76.7 to 81.9 m/sec) (comminuted fracture).

Studies of Relationships Between Skin Subcutaneous Flap and Bone Flap

There was no apparent relationship between both flaps. We have observed that bone fracture can occur without any skin wound. Most of the time, wound aspects are not similar, and it is unlikely to anticipate fracture on bone flap, by observation of the skin flap alone. Flap-by-flap dissection must be carried out to search bone fracture, and the pathologist must look for deep wounds, even if a superficial flap is woundless.

According to these lesions, described by flap-by-flap dissections, it is important to define a bullet-skin-bone entity as the interaction of the hybrid ammunition on skin with bone tissues. Variations in velocity simulate variation in distance ranges. Nevertheless, disparity of results does not allow any correlation between skin and bone lesions. Biomechanical properties of PMHS are different from living subjects. Transposition of results can not be performed without correcting factors.

With the hybrid projectile we tested, there was no correlation between distance range and lesions. This is a complete difference with what has been observed with homogeneous rubber projectiles (8,9). Therefore, the forensic pathologist has to be very careful when evaluating a distance range from lesions resulting from hybrid LLP.

References

1. Millar R, Rutherford W, Johnson S, Malhotra VJ. Injuries caused by rubber bullets: a report on 90 patients. Br J Surg 1975;62:480–6.
2. Schyma C, Schyma P. Possibilities for injuries caused by rubber bullets from the self-defense weapon MR-35 Punch [German]. Arch Kriminol 1997;200:87–94.
3. Wahl P, Schreyer N, Yersin B. Injury pattern of the Flash-Ball, a less-lethal weapon used for law enforcement: report of two cases and review of the literature. J Emerg Med 2006;31(3):325–30.
4. Jussila J, Normia P. International law and law enforcement firearms. Med Confl Surviv 2004;20(1):55–69.
5. Suyama J, Panagos PD, Sztajnkrycer MD, FitzGerald DJ, Barnes D. Injury patterns related to use of less-lethal weapons during a period of civil unrest. J Emerg Med 2003;25(2):219–27.
6. Luyer MD, Hoofwijk AG. Blunt liver trauma from bean bag ammunition. Eur J Trauma Emerg Surg 2008; doi: 10.1007/s00068-008-8059-z.
7. de Brito D, Challoner KR, Sehgal A, Mallon W. The injury pattern of a new law enforcement weapon: the police bean bag. Ann Emerg Med 2001;38(4):383–90.
8. Voiglio EJ, Frattini B, Dörrzapf JJ, Breteau J, Miras A, Caillot JL. Ballistic study of the SAPL GC27 gun: is it really ''nonlethal''? World J Surg 2004;28(4):402–5.
9. Bir CA, Stewart SJ, Wilhelm M. Skin penetration assessment of less lethal kinetic energy munitions. J Forensic Sci 2005;50(6):1426–9.

Additional information and reprint requests:
Humbert de Freminville, M.D., M.Sc.
UMRESTTE UMR T 9405
Université de Lyon
Université Claude Bernard Lyon 1
Faculté de Médecine Lyon-Est Claude Bernard
8, Rockefeller Ave
F-69373 Lyon
Cedex 08
France
E-mail: humbert.defreminville@orange.fr; hdefreminville@free.fr

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01439.x
Available online at: interscience.wiley.com

CASE REPORT

PATHOLOGY AND BIOLOGY

Audrey Farrugia,[1] *M.D.; Jean-sébastien Raul,*[1] *M.D., Ph.D.; Annie Géraut,*[1] *M.D.; and Bertrand Ludes,*[1] *M.D., Ph.D.*

Ricochet of a Bullet in the Spinal Canal: A Case Report and Review of the Literature on Bullet Migration

ABSTRACT: Ricochet of a bullet in the spinal canal is well known by neurosurgeons but relatively not a common event in usual medico-legal autopsy practice. This article presents a homicide case of a penetrating gunshot injury of the lumbar spine through the T12-L1 intervertebral foramen with active movement of the projectile within the spinal canal to the L5-S1 level. This case illustrates a bullet intradural and intramedullary active movement because of a ricochet of the body of T12 with active redirection of the path. In the current literature, different types of migration in caudal or cranial direction, intradural, or intramedullary are reported. If spontaneous migration of T10 to S1 seems to be more frequent, some authors reported a C1 to S2 migration. Such migration could be asymptomatic or induce neurological impairment. The medico-legal consequences of these migrations within the spinal canal are described.

KEYWORDS: forensic science, bullet, ricochet, migration, spinal canal, neurological impairment

Bullet migration through the spinal canal is a rare condition of penetrating missile injuries (1) well known by neurosurgeons. A review of the literature shows few cases with intradural migration of a bullet within the spinal canal (Table 1). In forensic literature, we found no previous report of a ricochet within the spinal canal. However, like intravascular foreign body embolism (2,3), the phenomenon of bullet migration within the spinal canal and the active movement of the projectile because of ricochet off internal body structure have medico-legal consequences, which are described here.

We present a gunshot homicide case with a bullet entry into the spinal canal at T12-L1 level and a final active movement to L5-S1 level because of a ricochet of the T12 vertebral body.

The different types and mechanisms of migration and the neurological impairment induced are discussed on the basis of the current literature.

Case Report

A 22-year-old man was shot with a 7.65-mm handgun on the left side of the chest in the sixth intercostal space. The bullet had previously passed through the arm of a first victim with an intramuscular trajectory. The 22-year-old man immediately fell backward after being shot and was transported to the nearest hospital where he died in the course of an emergency exploratory laparotomy and left thoracotomy. External examination of the victim revealed two surgical incisions of laparotomy and left thoracotomy. The entrance gunshot wound was located on the left side of the chest, about 10 cm to the left of the midline and 14 cm below the left nipple (Fig. 1). The wound was round, measured 5×5 mm, and was surrounded by a symmetrical margin of 3-mm-wide abraded skin with a discrete ecchymotic area. There was no soot, gunpowder residue, or stippling around the wound. No exit wound was observed. X-rays showed a metallic object located in the spine area at the level of L5-S1 corresponding to the projectile (Fig. 2).

During autopsy, we found that the bullet passed through the sixth left intercostal space, and then entered the left inferior pulmonary lobe, the left hepatic lobe, the stomach, the pancreas, and the left renal vessels. The renal lesions caused an extensive internal hemorrhage, which infiltrated the left psoas. After its intrathoracic and intra-abdominal pathway, the bullet passed through the posterior left peritoneum and entered into the spinal canal through the T12-L1 intervertebral foramen with small bone defect at the infero-lateral border of the T12 vertebral body. Dissection of the spine was performed via a posterior approach with classical laminectomy. We noted a rupture of the dura at T12 level and an epidural hemorrhage between T12 and the terminal extremity of the dura. After opening of the dura, we observed a laceration of the terminal spine and the cauda equina. The bullet was found among the cauda equina nerve roots at the L5-S1 level, about 110 cm above the heel level. The projectile was a slightly flattened 7.65-mm full metal jacket bullet (Fig. 2).

Toxicological investigations revealed 34.72 mM of alcohol in blood and were positive for ketamine that had been administered during surgical intervention. The cause of death was an extensive hemorrhage attributed to the hemorrhagic lesions of the stomach, the liver, and the left renal vessels.

Discussion

Intradural migration of bullets was first reported in 1916 (4) and later during World War I by Galem and Smith (5). They reported

[1]Institute of Legal Medicine, University of Strasbourg,11 rue Human 67085, Strasbourg Cedex, France.

Received 31 Mar. 2009; and in revised form 31 July 2009; accepted 8 Aug. 2009.

TABLE 1—*Reported cases of injuries with bullet migrating in the spinal canal.*

Authors	Entrance Wound	Bone Lesion	Migration	Neurological Status
Cagavi F et al. 2007 (7)	Right side of the abdomen	Fragmentation of the right pedicle at the L3 level	L3 to S2 *Intradural*	Paraparesis (proximally 1/5, distally 0/5), anesthesia below L3, loss of anal tone
Karim NO et al. 1986 (8)	Left lower quadrant of the abdomen	Not described	T11-T12 to L4-L5 *Subarachnoid space*	Initially no neurological impairment. After migration the patient developed left leg pain and left dorsiflexion weakness
Kafadar AM et al. 2006 (1)	Left upper quadrant of the abdomen	Defect at the anterolateral border of L1 corpus	L1 to S1-S2 *Intradural*	Paraplegia below L1
Rajan DK et al. 1997 (13)	Behind the right mastoid tip	No displaced fracture of the left posterior arch of C1	C1 to T6 and 3 years later to S2 *Intradural*	Mild weakness (4/5) of the proximal left upper extremity (at the time of injury or after migration)
Oktem IS et al. 1995 (16)	Left chest at the sixth intercostal space	Defect at the left T6 corpus	T6 to S2 *Subarachnoid space*	Paraplegia below T6
Tanguy A et al. 1982 (14)	In the neck at the C7 level	Not described	C7 to S1-S2 *Subarachnoid space*	No neurological impairment (at the time of injury or after migration)
Ben-Galim P and Reitman CA, 2008 (10)	Dorsally adjacent to the L3 vertebral body	Not described	L3 to S1 S1 to L3 and finally to T12 T12 to L5-S1 level with Trendelenburg position *Subarachnoid space*	No neurological impairment (at the time of injury or after migration)
Avci SB et al. 1995 (11)	Left lower quadrant of the abdomen	Not described	S1 to L4 *Intradural*	Initially mild hypoesthesia at S1 on the left side. After migration the patient developed loss of Achilles reflex and plantar flexion weakness
Gupta S and Senger RL, 1999 (12)	Right side of the lower chest	Not described	S1 to L3 *Subarachnoid space*	Initially no neurological impairment. After migration the patient developed foot drop on the right, partial foot drop on the left, urinary retention, decreased sensation in the S1-4 segments
Young WF et al. 1993 (18)	Head	Left maxilla and left orbit	Cranium to C5 over a course of 4 years	No neurological impairment during the migration
Arasil E and Tascioglu AO, 1982 (15)	Left occipital area	Not described	Cranium to C4 *Subarachnoid space*	Initially no neurological impairment After migration the patient developed Lhermitte's sign

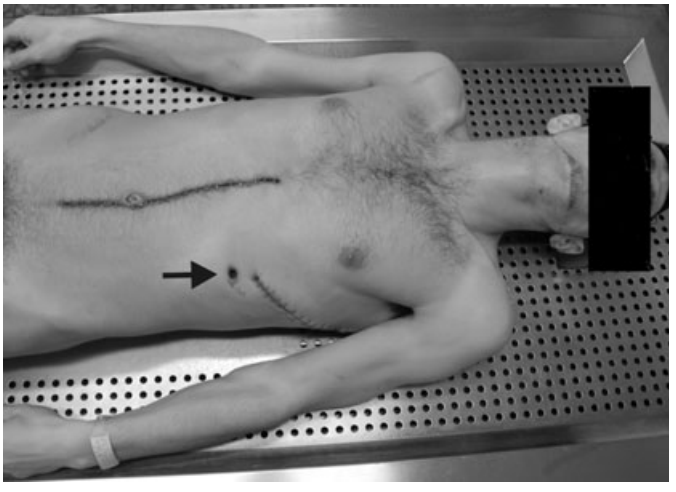


FIG. 1—*Lateral view of entrance wound in the left side of the chest and surgical incisions. Black arrow showing entrance wound with a peripheral margin of abraded skin.*

that 1 out of every 1500 cases of gunshot wounds to the head showed bullet migration on radiography. As in all other cases in which initial bullet trajectory is changed in the body, one of the most important goal of autopsy is to establish the bullet pathway. In this case, the autopsy revealed that the bullet pathway, from the entrance wound to T12, was left to right, backward, and slightly down. In the spinal canal, the change in trajectory of the bullet was because of a ricochet as a result of impact with the infero-lateral border of the T12 vertebral body. During its pathway within the spinal canal the bullet passed through the epidural space, the dura mater, the subdural space, the subarachnoid space, and the pia mater, and then induced a laceration of the terminal spine to finish its path in the middle of the cauda equina (Fig. 3).

This phenomenon of ricochet had previously been reported in a neurosurgical article (6). In this case the intradural and intramedullary movement was a direct result of the penetrating missile, which changed its direction and turned downward after hitting the anterior wall of the spinal canal. At autopsy, the authors described an entrance wound in the right side of the neck, a defect of the body of the C6 vertebra, at the point where the missile changed its direction by 80°, and a liquefaction of the spinal cord between the C6 and the T10 level where the bullet and its fragments were removed. In the case reported, the intracanal foreign body migration was not spontaneous motion because of the gravity (6).

In other cases reported in the literature, movements of bullets into the spinal canal occurred mainly because of gravitational forces (9), more frequently between T10 and S1-S2 (1,7,8).

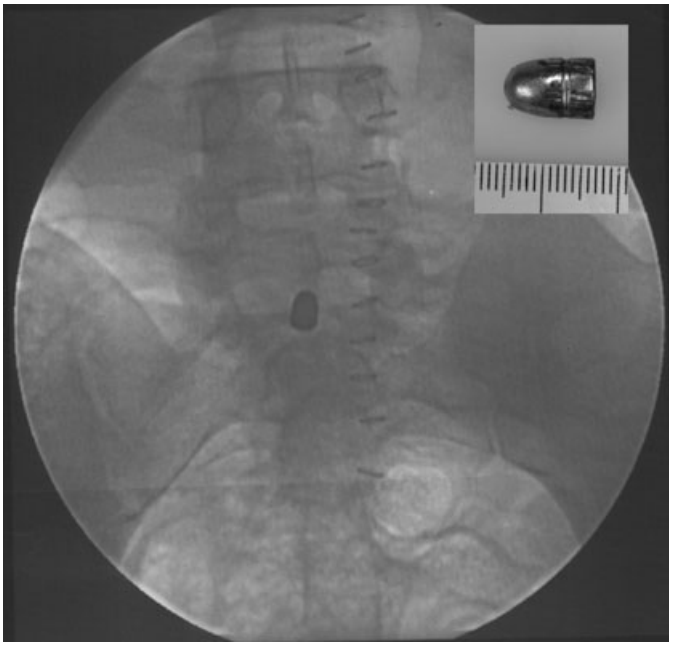


FIG. 2—*X-ray showing the bullet located at the L5-S1 level. The bullet recovered was a 7.65-mm full metal jacket bullet, slightly flattened.*

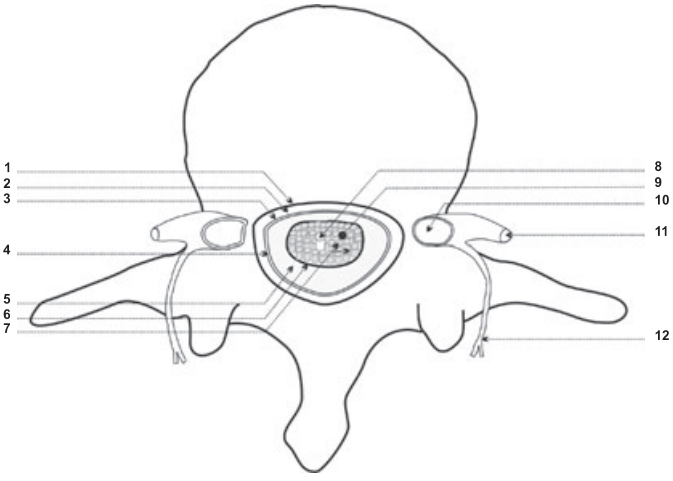


FIG. 3—*Anatomic diagram of the spinal cord in the lumbar region at L5 level (superior view): (1) spinal canal, (2) extradural space, (3) dura mater, (4) subdural space, (5) subarachnoid space, (6) pia mater, (7) cauda equina, (8) filum terminale, (9) final position of the bullet, (10) spinal ganglion, (11) ventral ramus, and (12) dorsal ramus of spinal nerve.*

Migration above the level of T10 is thought to be prevented by the narrow spinal canal at T10 (9). Forces such as coughing, swallowing, peristalsis, and blood flow may play a role in changing the direction of a bullet in the body, but the main factor resulting in migration within the spinal canal is the gravity (9).

According to Ben-Galim et al. (10), the missile susceptibility to the effects of gravity has been used to optimally position the bullet before the surgical procedure. Indeed, these authors described a case of bullet migration in subarachnoid space from T12 to L5-S1 level after maintaining the patient at bed rest in a reverse Trendelenburg position during 2 days. Atypical passive migrations in cranial direction (11,12) or from the cervical spinal canal to the sacrum (13,14) have also been described and can be explained by the effect of gravity related to the position of the body. It is important to understand that the passive movement of the projectile within the spinal canal could be an extradural migration in the epidural space or an intradural migration in subdural or subarachnoid space (Fig. 3). However, the intramedullary movement of the projectile in the spine needs some kinetic energy and constitutes consequently an active movement.

Concerning neurological impairment, it is important to note that spinal injuries from gunshot wounds are not universally associated with neurological impairment (Table 1).

Arasil et al. (15) reported a case of spontaneous intradural migration of a 7.65-mm caliber bullet from cranium to C4 level causing Lhermitte's sign, because of the compression of the spinal cord with the bullet. In other cases of bullet migration in lumbosacral spinal canal, namely T6 to S2, L1 to S1-S2, and L3 to S2, patients developed paraplegia or paraparesis and anesthesia below, respectively, T6, L1, and L3 level (1,7,16) (Table 1).

There is one report in which an intraspinal 9-mm caliber bullet lodged at L2-L3 level caused no neurological deficit (17). Although in this case the bullet has not migrated.

Finally, three previous articles described cases in which an intraspinal bullet migration from cranium to C5 (18), C7 to S1-S2 (14), and L3 to S1 (10) caused no neurological deficit. In one case (18) the bullet was fragmented and in two others the pathway was intradural but extramedullary.

With the disparity between the described cases and the victims' symptoms, we can wonder why in some cases patients developed neurological impairment and some others are asymptomatic. The explanation could be that when the bullet movements occurs below the L2 level, the bullet can potentially pass through the nerves roots of the cauda equina without causing damage. When the bullet movements occurs above the L2 level, the existence or not of neurological impairment could be induced by the caliber of the bullet, the potential fragmentation of the bullet, and the extra or intramedullary pathway of the bullet.

Conclusions

To the best of our knowledge, a ricochet within the spinal canal has not been previously described in the forensic literature. In cases reported in the literature, the movement of the bullet within the spinal canal is a migration that occurs generally in a caudal direction between T10 and S1-S2 because of the gravitational forces (9). Atypical passive intradural migrations in cranial direction or from cervical spine to the sacrum have also been previously reported. The case reported in this study illustrates that the internal trajectory of a projectile in the body cannot always be determined by a straight line connecting the entrance wound to the final resting place of the projectile or to the exit wound. Two phenomena can render that interpretation incorrect: internal bullet ricochet (active movement) and internal bullet migration (passive movement). Moreover, the neurological impairments are not systematic. The understanding of such phenomena is important for forensic scientists and death investigators to know whether the victim was or not able to move after sustaining such type of nonimmediately lethal gunshot injuries.

References

1. Kafadar AM, Kemerdere R, Isler C, Hanci M. Intradural migration of a bullet following spinal gunshot injury. Spinal Cord 2006;44(5):326–9.
2. Slobodan S, Slobodan N, Djordje A. Popliteal artery bullet embolism in a case of homicide: a case report and review of the tangible literature. Forensic Sci Int 2004;1:27–33.

3. Lucena JS, Romero C. Retrograde transthoracic venous bullet embolism. Report of a case following a single gunshot with multiple wounds in the left arm and chest. Forensic Sci Int 2002;3:269–72.
4. Villandre G, Morgan JD. Movement of foreign bodies in the brain. Radiol Electroh 1916;21:22–7.
5. Gamlen HE, Smith S. A study of interrelation between the radiography and surgery of gunshot wounds to the head. Br J Surg 1918;5:422–4.
6. Tekavcic I, Smrkolj VA. The path of a wounding missile along the spinal canal: a case report. Spine 1996;5:639–41.
7. Cagavi F, Kalayci M, Seckiner I, Cagavi Z, Gul S, Atasoy HT, et al. Migration of a bullet in the spinal canal. J Clin Neurosci 2007;14(1):74–6.
8. Karim NO, Nabors MW, Golocovsky M, Cooney FD. Spontaneous migration of a bullet in the spinal subarachnoid space causing delayed radicular symptoms. Neurosurgery 1986;1:97–100.
9. Ledgerwood AM. The wandering bullet. Surg Clin North Am 1977;57(1):97–109.
10. Ben-Galim P, Reitman CA. Intrathecal migratory foreign body without neurological deficit after a gunshot wound. Eur Spine J 2008;8(2):404–7.
11. Avci SB, Acikgoz B, Gundogdu S. Delayed neurological symptoms from the spontaneous migration of a bullet in the lumbosacral spinal canal. Case report. Paraplegia 1995;33(9):541–2.
12. Gupta S, Senger RL. Wandering intraspinal bullet. Br J Neurosurg 1999;13(6):606–7.
13. Rajan DK, Alcantara AL, Michael DB. Where's the bullet? A migration in two acts. J Trauma 1997;43(4):716–8.
14. Tanguy A, Chabannes J, Deubelle A, Vanneuville G, Dalens B. Intraspinal migration of a bullet with subsequent meningitis. A case report. J Bone Joint Surg Am 1982;64(8):1244–5.
15. Arasil E, Tascioglu AO. Spontaneous migration of an intracranial bullet to the cervical spinal canal causing Lhermitte's sign. Case report. J Neurosurg 1982;56(1):158–9.
16. Oktem IS, Selcuklu A, Kurtsoy A, Kavuncu IA, Pasaoglu A. Migration of bullet in the spinal canal: a case report. Surg Neurol 1995;44(6):548–50.
17. Jeffery JA, Borgstein R. Case report of a retained bullet in the lumbar spinal canal with preservation of cauda equina function. Injury 1998;29(9):724–6.
18. Young WF, Katz MR, Rosenwasser RH. Spontaneous migration of an intracranial bullet into the cervical canal. South Med J 1993;86(5):557–9.

Additional information and reprint requests:
Audrey Farrugia, M.D.
Forensic Scientist
Institute of Legal Medicine
11, rue Humann
67085 Strasbourg
Cedex
France
E-mail: audrey.farrugia@unistra.fr

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01407.x
Available online at: interscience.wiley.com

CASE REPORT

PATHOLOGY AND BIOLOGY

Roger W. Byard,[1,2] *M.D.; David Veldhoen*[3]*; Hilton Kobus,*[4] *Ph.D.; and Karen Heath,*[2] *M.B.B.S.*

''Murder–Suicide'' or ''Murder–Accident''? Difficulties with the Analysis of Cases

ABSTRACT: Homicide where a perpetrator is found dead adjacent to the victim usually represents murder–suicide. Two incidents are reported to demonstrate characteristic features in one, and alternative features in the other, that indicate differences in the manner of death. (i) A 37-year-old mother was found dead in a burnt out house with her two young sons in an adjacent bedroom. Deaths were due to incineration and inhalation of products of combustion. (ii) A 39-year-old woman was found stabbed to death in a burnt out house with her 39-year-old de facto partner deceased from the combined effects of incineration and inhalation of products of combustion. The first incident represented a typical murder–suicide, however, in the second incident, the perpetrator had tried to escape through a window and had then sought refuge in a bathroom under a running shower. Murder–accident rather than murder–suicide may therefore be a more accurate designation for such cases.

KEYWORDS: forensic science, murder–suicide, dyadic, accident, homicide, arson, gasoline

In situations where a murder has occurred and the perpetrator is found deceased at the scene, the most likely scenario is that this represents a so-called murder–suicide, homicide–suicide, or dyadic death (1–3). Under these circumstances, the victim has been killed and the perpetrator has committed suicide soon after. The motivations for the murders and the subsequent deaths of the perpetrators are extremely diverse, although a number of classical and idiosyncratic situations have been described (4–6). In the following report, two incidents are discussed to illustrate features that may be typical of murder–suicide and other features that may indicate a different scenario.

Incident Details

Incident 1

A report was received of a house fire in which a mother and her two young children had perished. Initially, the possibilities of accidental death or a triple homicide were considered; however, examination of the scene by fire investigators and police officers revealed evidence of accelerant use with a characteristic pattern of burning and a gasoline container within the house. Police interviews established that the three victims had been alone in the house at the time of the fire, suggesting that the children had been murdered by their mother who had then taken her own life in the fire. The autopsy findings are detailed below.

[1]Discipline of Pathology, The University of Adelaide, Frome Rd, Adelaide, SA, 5005, Australia.
[2]Forensic Science SA, 21 Divett Place, Adelaide, SA, 5000, Australia.
[3]Forensic Response Section, SAPOL, Adelaide, Australia.
[4]Flinders University, Adelaide, Australia.

Received 15 April 2009; and in revised form 7 June 2009; accepted 3 July 2009.

Case 1—An 18-month-old boy had been found dead in his bed in a burnt out house. The body smelt strongly of gasoline, and radiologic examination did not reveal injuries or projectiles. Extensive charring of the skin of the head, face, limbs, and torso was noted. No bony injuries or organic diseases were found. No subgaleal bruising was noted to suggest blunt cranial trauma. A small amount of soot was present in the airways. Toxicologic evaluation of blood and tissues revealed a carboxyhemoglobin level of 6%, no cyanide, with volatiles in headspace analyses in keeping with gasoline exposure. Death was attributed to incineration with the low level of carboxyhemoglobin attributed to rapid combustion because of the use of an accelerant. Although the possibility of prior death from suffocation could not be excluded, the strong odour of gasoline on the body suggested that accelerant had been poured onto the victim placing him at the center of the conflagration.

Case 2—A 4-year-old brother was found dead in the same bedroom. The body did not smell of gasoline, and radiologic examination did not reveal injuries or projectiles. The body was coated in soot with irregular areas of superficial burning of the face and limbs. There was extensive soot staining of the mouth, nose, teeth, and upper airways, and the skin and internal organs were cherry pink in color in keeping with carbon monoxide toxicity. No bony injuries or organic diseases were found. No subgaleal bruising was noted to suggest blunt cranial trauma. Toxicologic evaluation of blood and tissues revealed a carboxyhemoglobin level of 81%, a cyanide level of 0.15 mg/L, with volatiles in headspace analyses in keeping with gasoline exposure. Death was attributed to inhalation of products of combustion.

Case 3—The 37-year-old mother of the two boys was found dead in a bathroom in the center of the house. There was a history

of depression and no attempt had been made to escape. Extensive charring of the skin of the head, face, limbs, and torso was noted. No bony injuries or organic diseases were found. A small amount of soot was present in the airways. Toxicologic evaluation of blood and tissues revealed a mildly raised carboxyhemoglobin level of 19%, no cyanide, with volatiles in headspace analyses in keeping with gasoline exposure. Death was attributed to the combined effects of inhalation of products of combustion and incineration.

Incident 2

A report was received of a house fire in which the bodies of a woman and her de facto partner were found. The female victim had been stabbed. Examination of the scene by fire investigators and police officers revealed evidence of accelerant use with a gasoline container beside the female victim. Police interviews established that the victims had been alone in the house at the time of the fire. At the time of the initial police investigation, a double homicide was considered most likely. The autopsy findings are detailed below:

Case 1—A 39-year-old woman was found dead lying face down on the floor of a bedroom in a burnt out house. Extensive charring of the head, neck, and limbs was noted. In addition, 41 stab wounds to the chest, abdomen, and right thigh were identified, with stab wounds to the heart, lungs, liver, kidneys, spleen, and major vessels. No organic diseases were identified. No soot was present in the airways. Toxicologic evaluation of blood and tissues revealed a carboxyhemoglobin level of 8% with no other significant findings. No volatiles were identified on headspace analyses. Death was attributed to multiple stab wounds to the chest and abdomen and had occurred before the fire had started.

Case 2—The female victim's 39-year-old male partner was found dead lying face down on the floor of a shower cubicle, with the taps turned on, in an en suite bathroom of the bedroom. Superficial burns of the head, neck, and trunk were present, along with charring of the limbs. Irregular superficial incised wounds of the left forearm and palm were also present. No organic diseases were identified. Soot was present in the upper airways, and toxicologic evaluation of blood and tissues revealed a carboxyhemoglobin level of 46% and a cyanide level of 0.5 mg/L. In addition, there was a blood alcohol level of 0.027% with therapeutic concentrations of paracetamol and codeine. No volatiles were identified on headspace analyses. Death was attributed to the combined effects of inhalation of products of combustion and incineration.

Further information from the scene revealed that the window of the bathroom where the male victim had been found had been broken outwards suggesting that he had attempted to escape through this. This would account for the superficial incised wounds of the forearm and palm that were not in the usual position of suicidal wounds.

Discussion

When a perpetrator is found dead at a scene with the body of a victim, and there is no evidence of the involvement of a third party, the assumption is usually made that the case represents a murder–suicide where the perpetrator has intentionally taken his or her life after killing the other person(s). In the cases reported, this was one of the initial assumptions, along with multiple homicides; however, closer analysis of the circumstances revealed features in the second incident that suggest an alternative scenario.

Murder–suicides are rare compared to homicides where the perpetrator survives and have the lowest rates in communities where the homicide rate is high (7). For example, in Denmark, which has a low homicide rate, 42% of homicide cases are murder–suicides, compared to the United States with very high homicide rates where the percentage is a mere 4% (8). Rates range from 0.05/100,000 of the population in New Zealand to 0.21/100,000 in Canada (6).

A variety of classifications have been used for murder–suicides with Marzuk et al. (9) proposing three broad groups of spousal, familial, and extrafamilial. Spousal murder–suicides generally separate into those involving physically ill and aging couples, and those of younger, possessive, male perpetrators who have often been rejected by a lover or spouse. It was initially considered that the cases from the second incident fell into the latter category. Familial murder suicides generally involve a parent killing their children and then themselves. This was thought to be the scenario in the first incident. Extrafamilial murder–suicides involve a range of situations including disgruntled former employees, and those with pseudocommando or cult features (9–14). Others have used more complex classifications to delineate more clearly the relationship of the victim to the perpetrator and factors that may have precipitated the event (15).

In the reported incidents, gasoline was being used as an accelerant to assist with the setting of fires, as it is volatile and highly flammable. Combustion of hydrocarbon fuels of this nature is an oxidation reaction and therefore requires an oxygen source from the surrounding air. The liquid fuel itself will not burn as it is the vapour mixed with the air above the surface that ignites. When gasoline is poured onto the floor of a premise in cases of arson, a large surface area of the fuel is created enhancing evaporation and a widespread flammable air/fuel atmosphere can result. The flammable properties of the mixture are determined by the ratio of fuel vapour to air. When this mixture is within certain limits, the speed of combustion is so fast that an extremely rapidly developing fire or even an explosion may result. The limits for gasoline vapour/air explosions to occur are when the gasoline concentration lies between 1.1% and 6% (16). When the concentration is above these limits (a rich mixture) or below (lean mixture), no ignition results.

While the death of the perpetrator in the first incident was considered a murder–suicide, the death of the perpetrator in the second incident may have resulted from an underestimation of the potential intensity and rapidity of spread of an accelerant-enhanced indoor fire. It appears that the fire that was being used to disguise a homicide had resulted in the perpetrator being trapped with his victim, and dying as a result of smoke inhalation and burns. Evidence supporting the unintentional nature of the lethal event for the perpetrator included his attempts to escape and his seeking refuge in a bathroom under a running shower. Given the emotional stress of the circumstances, it is perhaps not surprising that miscalculation had occurred, with failure to plan a clear escape route. An absence of self-inflicted injuries, such as typical multiple parallel incised wounds of the wrists, and the lack of excessive or lethal levels of drugs on toxicological screening were also in keeping with the accidental nature of the death, although it remains a possibility that an initial decision to commit suicide by self-immolation had been reversed. In this case, however, death would still be classified as accidental.

In conclusion, cases of apparent murder–suicide where an accelerant-enhanced fire has been started may represent a murder followed by the accidental death of the assailant while attempting to disguise the crime by burning the victim's body and surroundings. If this is the case, then the different manner of death requires

separation of the two entities, as the psychological profile of the perpetrator may be different in an aborted suicide attempt or an accidental death because of miscalculation, compared to an intentional death from a deliberate act.

Acknowledgment

We thank the South Australian State Coroner, Mr. Mark Johns, for permission to report selected details of these cases.

References

1. Milroy C. Murder suicide. In: Payne-James J, Byard RW, Corey TS, Henderson C, editors. Encyclopedia of forensic and legal medicine, Vol 3. Amsterdam: Elsevier/Academic Press, 2005;358–61.
2. Byard RW. Murder-suicide—an overview. In: Tsokos M, editor. Forensic pathology reviews, Vol 3. New York, NY: Humana Press, 2005;337–47.
3. Copeland AR. Dyadic death revisited. J Forensic Sci Soc 1985;25:181–8.
4. Milroy CM. Reasons for homicide and suicide in episodes of dyadic death in Yorkshire and Humberside. Med Sci Law 1995;35:213–7.
5. Byard RW, Knight D, James RA, Gilbert J. Murder-suicides involving children—a 29 year study. Am J Forensic Med Pathol 1999;20:323–7.
6. Milroy CM. The epidemiology of homicide-suicide (dyadic death). Forensic Sci Int 1995;71:117–22.
7. Coid J. The epidemiology of abnormal homicide and murder followed by suicide. Psychol Med 1983;13:855–60.
8. Felthous AR, Hempel A. Combined homicide-suicide: a review. J Forensic Sci 1995;40:846–57.
9. Marzuk PM, Tardiff K, Hirsch CS. The epidemiology of murder-suicide. JAMA 1992;267:3179–83.
10. Milroy CM, Dratsas M, Ranson DL. Homicide-suicide in Victoria, Australia. Am J Forensic Med Pathol 1997;18:369–73.
11. Palermo GB, Smith MB, Jentzen JM, Henry TE, Konicek PJ, Peterson GF, et al. Murder-suicide of the jealous paranoia type. A multicenter statistical pilot study. Am J Forensic Med Pathol 1997;18:374–83.
12. Milroy CM. Homicide followed by suicide (dyadic death) in Yorkshire and Humberside. Med Sci Law 1993;33:167–71.
13. Hannah S, Turf E, Fierro M. Murder-suicide in central Virginia. Am J Forensic Med Pathol 1998;19:275–83.
14. Campanelli C, Gilson T. Murder-suicide in New Hampshire, 1995–2000. Am J Forensic Med Pathol 2002;23:248–51.
15. Hanzlick R, Koponen M. Murder-suicide in Fulton County, Georgia: comparison with a recent report and proposed typology. Am J Forensic Med Pathol 1994;15:168–73.
16. Kirk PL. Fire investigation, including fire-related phenomena: arson, explosion, asphyxiation. New York, NY: John Wiley and Sons, 1969.

Additional information and reprint requests:
Roger W. Byard, M.D.
Discipline of Pathology
Level 3 Medical School North Building
The University of Adelaide, Frome Road
Adelaide 5005
Australia
E-mail: byard.roger@saugov.sa.gov.au

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01397.x
Available online at: interscience.wiley.com

CASE REPORT

PATHOLOGY AND BIOLOGY

Efthimios Sivridis,[1] *M.D., Ph.D.; Pavlos Pavlidis,*[2] *M.D.; Charilaos Stamos,*[1] *M.D.; and Alexandra Giatromanolaki,*[1] *M.D.*

Sudden Death After Myocardial Infarction in a High-School Athlete

ABSTRACT: This is an unusual case of ischemic heart disease occurring in a young female athlete, aged 14 years, in whom almost the entire posterior wall of the left ventricle was affected. The patient collapsed and died suddenly after a vigorous physical activity while resting at home. The ischemic lesion was apparently owing to right orifice stenosis in association with a small hypoplastic right coronary artery.

KEYWORDS: forensic science, myocardial infarction, orifice stenosis, coronary artery hypoplasia, young athlete

A 14-year-old girl, who was a competitive high-school athlete, collapsed and died after severe exertion on the athletic field while resting at home. There was no history of heart disease, systemic hypertension, high cholesterol levels, or drug intake including cocaine, ephedrine, and anabolic steroids. None in the family had an unexplained death or heart attack at a young age.

At postmortem, the heart was of normal weight (224 g); yet, on cut sections, there was thinning of ventricular wall and a massive myocardial infarct, recognized as a dense white scar tissue, which involved almost the entire posterior wall of the left ventricle and the posterior 1/3 of the interventricular septum. The coronary arteries were patent and free of atheroma. However, the right coronary artery was short and small in caliber, and it gave off no branches to the posterior left ventricular wall. It was accompanied by a small slit-like right orifice. The left main and left anterior descending and circumflex coronary arteries were normal, as were their corresponding ostia and the cardiac valves. No anatomical abnormality was identified at necropsy in an organ other than the heart. The lungs were heavy and edematous with copious fluid running from their cut sections. The liver showed a characteristic "nutmeg" pattern. Meticulous inspection of other internal organs, including brain, kidneys, spleen, and intestine, was unremarkable.

Histological examination confirmed the presence of an old transmural myocardial infarct, with extensive areas of interstitial fibrosis involving the posterior left ventricular wall, the interventricular septum, and the papillary muscles. There was no disorganization of cardiac muscle cells and no evidence of chronic myocarditis. Both lungs showed severe pulmonary edema and intra-alveolar hemorrhage. The liver showed passive congestion.

Toxicologic tests for levels of alcohol, barbiturates, and amphetamines were performed, and they were all normal.

[1]Department of Pathology, Democritus University of Thrace, Medical School, Alexandroupolis, Greece.
[2]Department of Forensic Medicine, Democritus University of Thrace, Medical School, Alexandroupolis, Greece.

Received 3 May 2009; and in revised form 10 June 2009; accepted 26 June 2009.

Discussion

There have been sporadic reports describing the occurrence of sudden and unexpected death in young athletes during or just after severe exertion on the athletic field (1–4). Most of such fatal events are because of an underlying structural cardiovascular disease, usually hypertrophic cardiomyopathy (1). Other less frequent causes of unexpected death include anomalous origin of the left coronary artery from the anterior sinus of Valsalva, idiopathic concentric left ventricular hypertrophy, and ruptured aorta (1). Additional risk factors for young athletes comprise mitral valve prolapse, acute myocarditis, and prolonged QT interval syndrome (1). It is of interest that all the above cardiovascular abnormalities are, by and large, unsuspected during life.

Myocardial infarction as a cause of sudden death in young athletes and high-school children is rare but does occur. Recently,

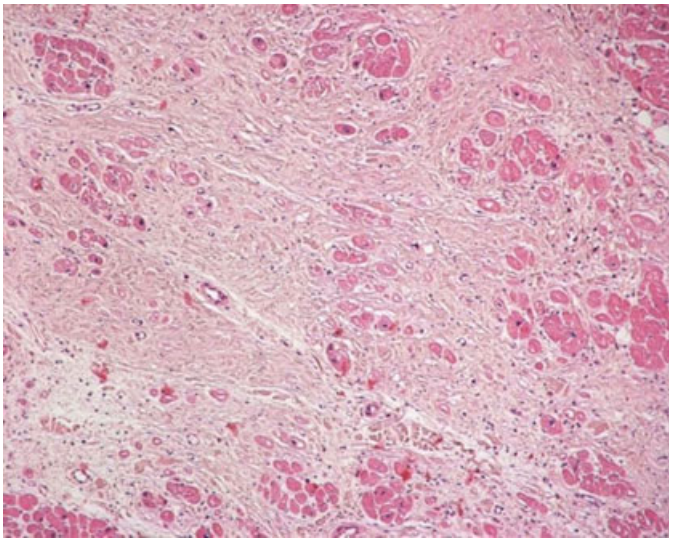


FIG. 1—*Exercise-induced myocardial infarction in a background of coronary artery hypoplasia in a high-school athlete (H-E).*

cocaine, ephedrine, and anabolic-androgenic steroids were correlated with ischemic heart disease in healthy, physically active young individuals, to promote athletic performance (2,3,5). Forte et al. reported the case of a young athlete with no risk factors for cardiovascular disease who experienced a myocardial infarction during the period in which he used an ephedrine-rich dietary supplement (2). Another case of acute myocardial infarction in a young athlete, which was also provoked by ephedrine abuse, has been described by Kranjec et al. (3). Agirbasli et al. reported a young male athlete who had an acute myocardial infarction after strenuous exercise (6).

In our case, cardiac disability in the high-school female athlete resulted from reduced coronary artery perfusion because of right orifice stenosis and right coronary artery hypoplasia. The long-standing ischemia caused a diffuse myocardial fibrosis with decreased muscular efficiency and consequently gradual failure of both left and right ventricles. Back-pressure effects of right ventricular failure was seen in the lungs as pulmonary edema and intra-alveolar hemorrhage, and as passive congestion in the liver. There was not obstructive hypertrophic cardiomyopathy and no history of drug usage. Coronary artery anomalies, including hypoplasia and orifice stenosis, are known causes of myocardial ischemia and sudden death in children and young adults (1). Patanè et al. presented recently a case of an acute myocardial infarction with diminutive right coronary artery, but this condition was associated with obstructive hypertrophic cardiomyopathy (4).

This case illustrates the need for early identification of asymptomatic cardiovascular disorders in competitive young and high-school athletes, through systematic preparticipation screening, to reduce the risk of sudden cardiac death.

References

1. Maron BJ, Roberts WC, McAllister HA, Rosing DR, Epstein SE. Sudden death in young athletes. Circulation 1980;62:218–9.
2. Forte RY, Precoma-Neto D, Chiminacio Neto N, Maia F, Faria-Neto JR. Myocardial infarction associated with the use of a dietary supplement rich in ephedrine in a young athlete. Arq Bras Cardiol 2006;87:e179–1.
3. Kranjec I, Cerne A, Noc M. Ephedrine-induced acute myocardial infarction in a young athlete: a case of thrombus management. Angiology 2009;60:254–8.
4. Patanè S, Marte F, Di Bella G, Chiribiri A. Acute myocardial infarction with diminutive right coronary artery and obstructive hypertrophic cardiomyopathy without significant coronary stenoses. Int J Cardiol 2009;135(3):e73–5.
5. Welder AA, Melchert RB. Cardiotoxic effects of cocaine and anabolic-androgenic steroids in the athlete. J Pharmacol Toxicol Methods 1993;29:61–8.
6. Agirbasli M, Martin GS, Stout JB, Jennings HS III, Lea JW IV, Dixon JH Jr. Myocardial bridge as a cause of thrombus formation and myocardial infarction in a young athlete. Clin Cardiol 1997;20:1032–6.

Additional information and reprint requests:
Alexandra Giatromanolaki, M.D.
Democritus University of Thrace
Department of Pathology
Alexandroupolis 68100
Greece
E-mail: agiatrom@med.duth.gr

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01413.x
Available online at: interscience.wiley.com

CASE REPORT

PATHOLOGY AND BIOLOGY

Francesco Ventura,[1] *M.D., Ph.D.; Alessandro Bonsignore,*[1] *M.D.; Raffaella Gentile,*[2] *M.D.; and Francesco De Stefano,*[1] *M.D.*

Two Fatal Cases of Hidden Pneumonia in Young People*

ABSTRACT: Acute respiratory distress syndrome (ARDS) is a severe lung disease characterized by inflammation of the lung parenchyma leading to impaired gas exchange. This condition is often lethal, usually requiring mechanical ventilation and admission to an intensive care unit. We present two fatal cases of hidden pneumonia in young people and discuss the pathophysiological mechanism of ARDS with reference to the histological pattern. A complete forensic approach by means of autopsy and histological, immunohistochemical, and microbiological, examination was carried out. In both cases the cause of death was cardio-respiratory failure following an acute bilateral pneumonia with diffuse alveolar damage and ARDS associated with sepsis and disseminated intravascular coagulation. Our cases suggest on one side the importance of an early diagnosis to avoid unexpected death while on the other that the diagnosis of ARDS has to be confirmed on the basis of a careful postmortem examination and a complete microscopy and microbiological study.

KEYWORDS: forensic science, forensic pathology, hidden pneumonia, adult respiratory distress syndrome, diffuse alveolar damage, sudden death, young people

Acute respiratory distress syndrome (ARDS) is a severe lung disease characterized by inflammation of the lung parenchyma leading to impaired gas exchange with concomitant systemic release of inflammatory mediators by local epithelial and endothelial cells, causing inflammation, hypoxemia resulting often in multiple organ failure (MOF), and disseminate intravascular coagulation (DIC) (1). This condition is often lethal, usually requiring mechanical ventilation and admission to an intensive care unit (2).

Physiopathologically when the endothelium of lung capillaries and the alveolar epithelium are damaged, plasma and blood flood the interstitial and intra-alveolar spaces. Such a change implies decreased lung compliance, pulmonary hypertension, reduced functional capacity, compromised ventilation/perfusion ratio, and hypoxemia (3).

Acute respiratory distress syndrome can occur within 24–48 h of an injury or attack of acute illness. In such a case the patient usually presents with shortness of breath and tachypnea, usually associated with hypoxemia, petechiae in the axillae, and neurologic abnormalities such as mental confusion (4).

Typical histological presentation involves diffuse alveolar damage (DAD) and hyaline membrane formation in alveolar walls (5). Hyaline membranes, especially, as a result of the acute inflammatory processes in the alveolar compartment (6) is the histological hallmark of ARDS.

[1]Department of Legal and Forensic Medicine, University of Genova, Via de Toni 12, 16132 Genova, Italy.
[2]Department of Clinical Pathology, San Martino Hospital, Piazza R. Benzi 10, 16132 Genova, Italy.
*Presented at the 61st Annual Meeting of the American Academy of Forensic Sciences, February 16–21, 2009 in Denver, CO.

Received 30 Mar. 2009; and in revised form 10 July 2009; accepted 12 July 2009.

If the underlying disease or injurious factor is not removed, the amount of inflammatory mediators released by the lungs in ARDS may result in a systemic inflammatory response syndrome (or sepsis if there is lung infection) (7). The evolution toward shock and/or MOF follows the same pathophysiological path of sepsis (8).

It is estimated that ARDS is caused by septic shock—characterized by leukocytosis or leukopenia, fever, hypotension, and the identification of a potential source of systemic infection with positive blood cultures for pathogenous agents—in more than 30% of cases (9).

Pneumonia is thought to be the most common lung disease leading to ARDS as it determines a direct lung injury in the immunocompetent host (10).

In cases of severe ARDS the survival rate is 50% with appropriate and early treatment, but if the ARDS-induced severe hypoxemia is not recognized and treated or if the disease reaches medical attention only in the terminal phase then cardio-respiratory arrest occurs in more than 90% of patients (11).

We present two fatal cases of hidden pneumonia in young people who died within a few hours. The clinical presentation, the radiological and laboratory findings in one case, and the postmortem examination with histological, immunohistochemical, and microbiological exams in both cases, led us to conclude for an acute cardio-respiratory failure secondary to bilateral pneumonia with DAD and consequently ARDS associated with sepsis and DIC. The features of the disease are discussed with reference to the histological and immunohistochemical evaluation.

Case 1

A 29-year-old man was found dead at home by his girlfriend who was sleeping with him. The night before he went out with

his friends and came back home late. His friends reported that nothing "strange" happened during the evening he spent with them. The day after, in the afternoon, when he should wake up his girlfriend saw the presence of foam around his mouth and nose and when she tried to wake him up he did not respond. So she called the emergency services who could do nothing but declare him dead.

He took psychodrugs and was known to be a drug addict and a heavy drinker. Family history was reported negative for sudden death. Death scene investigation was unremarkable.

Autopsy Findings

A complete postmortem examination was performed 4 days after death. External examination did not show any visible sign of injury. The internal examination revealed polyvisceral stasis, diffuse microthrombosis, cerebral and pulmonary edema. Free citrine liquid was found on both sides of the pleural cavities. A marked lung congestion and the release of foamy material were bilaterally observed. "Hydrostatic docimasia" for large and small fragments was bilaterally positive in all fields. Also known as "the flotation test," or "the lung test," this old test is still in use to check if there are areas of increased density within the adult lung parenchyma. In these cases lung specimens, being not inflated with air, do not float. Such is the case of pneumonia (12).

Histological Studies and Findings

The microscopic histological study, performed using formalin-fixed paraffin-embedded tissue sectioned at 4 μm and stained with hematoxylin-eosin (H&E), revealed the typical findings of DAD: alveolar septa mildly thickened by edema and capillary congestion, alveolar edema, hyaline membranes lining the denuded alveolar walls, hyperplastic type II pneumocytes, alveolar infiltrates of polymorphonuclear neutrophilic leukocytes, pigmented macrophages, monocytes and plasma cells (Fig. 1*A*), fibrin thrombi in small arteries. In some fields, numerous endoalveolar erythrocytes were also observed. Bronchial walls presented epithelial denudation, inside the lumen there were infiltrates of leukocytes, mostly neutrophils, and a moderate quote of eosinophilic amorphous material.

All these findings were suggestive for a typical DAD in the early exudative phase, confirmed by the positive results to immunohistochemical dye for surfactant apoprotein (PE-10) that outlines hyperplastic type II pneumocytes (13). Fungal infections were not found on slides by Grocott staining. Gram staining did not give evidence for bacterial colonies. The examination of other histological samples was unremarkable.

The lung samples were also examined under a confocal laser scanning microscope (14), and a three-dimensional reconstruction was performed (Fig. 2).

Microbiological Studies and Findings

Additional microbiological tests (15) to identify possible pathogenous agents were carried out through isolation of nucleic acids from formalin-fixed paraffin-embedded tissue sections. To control the course of extraction and check for PCR inhibitors, a fragment of the *Homo sapiens* beta-globin gene was amplified. The purified DNA sample was negative for all bacterial cultures. The positive result for beta-globin demonstrated that the DNA extraction procedure was efficient in extracting amplifiable DNA from the sample.

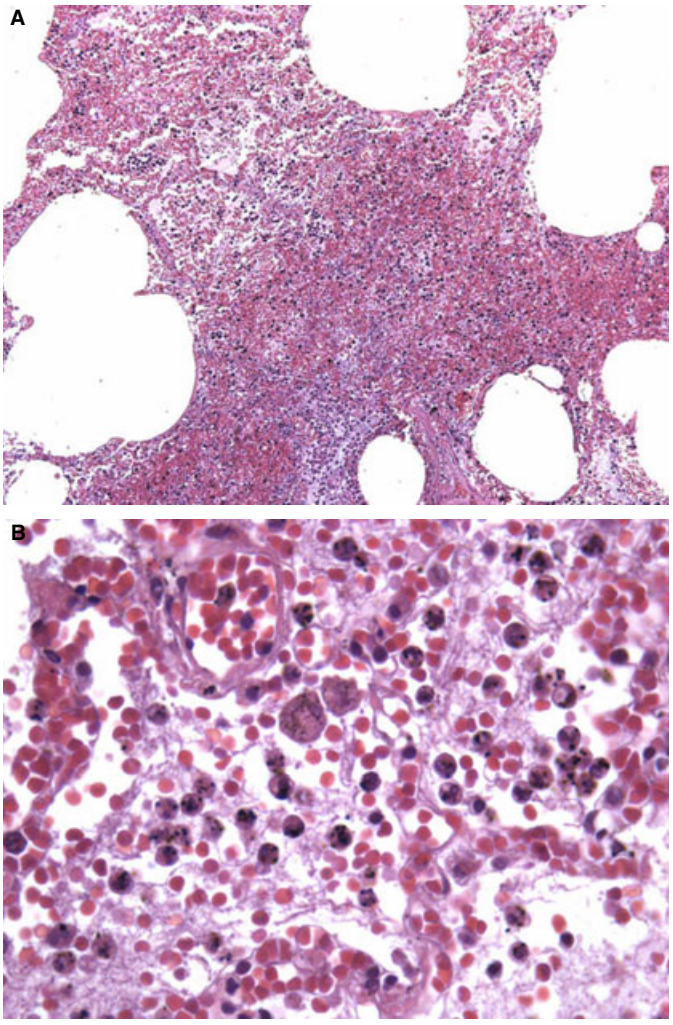


FIG. 1—*Alveolar infiltrates of polymorphonuclear neutrophilic leukocytes, pigmented macrophages, monocytes and plasma cells, and hyaline membranes (Hematoxylin and eosin). Case 1 (A) and case 2 (B).*

Toxicology was negative for drugs and alcohol. Thus, viral infection was a diagnosis of exclusion, according with recent literature which reports a prevalence of viral etiologies in community-acquired pneumonia up to 9% (16) and an extremely high incidence of lung injury and ARDS arising from coronavirus and avian influenza virus infection (17,18).

Case 2

A 31-year-old previously healthy man presented to the hospital with a 12-h history of sore throat, fever, and cough. The clinical prodromes were followed by the acute onset of increasing shortness of breath quickly progressing in acute respiratory failure with hemoptysis. Chest X-ray demonstrated bilateral diffuse airspace opacification; the high-resolution CT (HRCT) confirmed the presence of bilateral, symmetric diffuse ground-glass attenuation associated with liquid in pleural cavities. The patient was admitted to the intensive care unit with severe leukopenia, but he got worse and after few hours died. Two postmortem blood cultures were positive for group A beta-hemolytic Streptococcus which is well known for causing invasive disease leading to death even though diagnosis is not always made in life, as in this case (19). No other pathogenous agents were present.

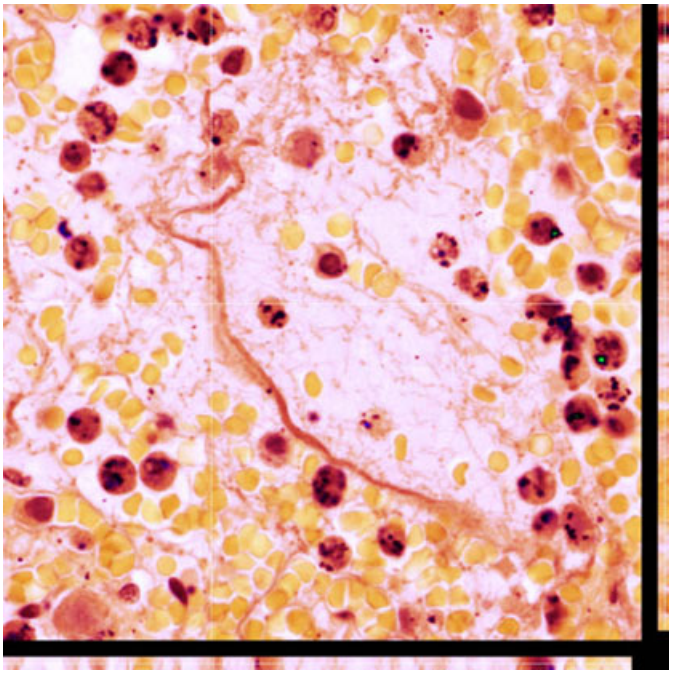


FIG. 2—*Three-dimensional reconstruction of hyaline membranes examined under a confocal laser scanning microscope.*

Autopsy Findings

An autopsy was performed within 48 h after death. External examination was irrelevant. Internal examination revealed an increased consistency and weight of the lungs (1070 g the left and 1100 g the right respectively) with positive hydrostatic docimasia in all fields and intense congestion which was ascribed to a bilateral pneumonia. The examination of other organs was unremarkable except for intense polyvisceral stasis.

Histological Studies and Findings

The histological examination of lung specimens (one sample per lobe and more samples in increased thickening pulmonary areas as common practice), performed by using the same method described earlier for the previous case, showed alveolar septa that mildly thickened by edema and capillary congestion, alveolar edema, hyaline membranes coating alveolar septal surfaces, flattened pneumocytes, alveolar infiltrates of polymorphonuclear neutrophilic leukocytes, pigmented macrophages, monocytes and plasma cells, fibrin thrombi in small arteries (Fig. 1*B*). All these findings suggested a typical DAD, confirmed by positive results to immunohistochemical dye for surfactant apoprotein (PE-10) (data not shown) (13).

In the kidneys was found a thrombotic microangiopathy compatible with DIC. The lung samples were examined under a confocal laser scanning microscope, and a three-dimensional reconstruction was performed (14).

Results and Discussion

The silent (case 1) and the paucisymptomatic (case 2) presentations, and the histological and immunohistochemical findings led us to the diagnosis of ARDS supporting the conclusion that both were affected by a quite rare type of pulmonitis definable as hidden pneumonia.

Acute respiratory distress syndrome is a pathological entity arising from multiple pulmonary or extrapulmonary causes (20). Generally, patients with ARDS report a short prodromal illness characterized by few symptoms like fever and cough, followed by the acute onset of progressive shortness of breath which rapidly evolves to respiratory failure (21). Chest radiographs typically show bilateral diffuse airspaces opacifications (22). Chest HRTC scans are significative for bilateral ground-glass attenuation (23). The histological features of ARDS, investigated from open lung biopsies or autopsies, are those of DAD, a nonspecific pattern of acute lung injury (24).

Acute respiratory distress syndrome has a poor prognosis, with reported mortality rates still appearing to be higher than 50% (25).

The first case concerns a 29-year-old man found lifeless at home by his girlfriend. At autopsy polyvisceral stasis, diffuse microthrombosis, free citrine liquid on both sides of the pleural cavities, a marked lung congestion, and the release of foamy material were found. Hydrostatic docimasia for large and small fragments was bilaterally positive in all fields. The histological evaluation of lungs samples stained with H&E, also examined under a confocal laser scanning microscope, gave evidence of a pattern of DAD. The presence of hyperplastic type II pneumocytes and hyaline membranes was confirmed by the positive reaction of the immunohistochemical dye for surfactant apoprotein (PE-10). Additional tests were carried out to identify possible pathogenous agents through microbiological studies but all the cultures showed no bacterial growth. Toxicology was negative for drugs and alcohol.

The second case involves a 31-year-old previously healthy man who presented, after a 12-h history of sore throat, fever, and cough, an acute onset of increasing shortness of breath rapidly progressing in acute respiratory failure with hemoptysis. Chest X-ray and HRCT showed the typical pattern of ARDS, with bilateral, symmetric, diffuse ground-glass attenuation. Despite admittance to the intensive care unit, the patient died after few hours. Two blood cultures were positive for group A beta-hemolytic Streptococcus. The macroscopical and histological patterns were similar to that of case 1.

In both cases the cause of death was attributed to an acute cardio-respiratory failure secondary to acute bilateral pneumonia and consequently ARDS, sepsis, and DIC.

These cases demonstrate how ARDS can rapidly lead to death in young patients that can generally be successfully treated in case of pneumonia.

In the first case the postmortem diagnosis of ARDS and sepsis with DIC (26) was made exclusively on the basis of a careful postmortem examination and a complete histological study.

Therefore, the authors underline that forensic pathological procedures should be applied in all cases of sudden death using systematic practical investigations to find the cause of death, more so in fatal cases involving young people. Just in this way it is possible to perform an adequate differential diagnosis when sudden cardiac death is more likely to be expected because of the young age of the patient.

The second case, which has attracted the medicolegal interest because of medical liability profiles that were assumed as fault for doctors, suggests that clinicians should be suspicious of all community-acquired pneumonia (27), especially in young people, because rigorous diagnosis as well as early and appropriate therapy is mandatory to avoid unexpected death (28).

Particularly, from a forensic point of view, in such cases the authors suggest the importance of taking postmortem bacterial and viral cultures.

Last, the forensic community should not forget the role played by ARDS as a potential cause of sudden and unexpected death in previously healthy young people.

Acknowledgments

The authors thank Prof. Vittorio Fineschi for his help in histological and immunohistochemical studies. They also thank Margherita Neri, M.D. and Irene Riezzo, M.D. for their excellent technical assistance in confocal microscopy. Finally the authors thank Claudio Giacomazzi, M.D. for his help in microbiological studies.

References

1. Hardaway RM, Vasquez Y. A shock toxin that produces disseminated intravascular coagulation and multiple organ failure. Am J Med Sci 2001;322(4):222–8.
2. Malhotra A. Low-tidal-volume ventilation in the acute respiratory distress syndrome. N Engl J Med 2007;357(11):1113–20.
3. Bernard G, Artigas A, Brigham K, Carlet J, Falke K, Hudson L, et al. The American-European Consensus Conference on ARDS. Definitions, mechanisms, relevant outcomes, and clinical trial coordination. Am J Respir Crit Care Med 1994;149(3):818–24.
4. Irwin RS, Rippe JM. Irwin and Rippe's intensive care medicine, 5th edn. Baltimore, MD: Lippincott Williams & Wilkins, 2003.
5. Cotran RS, Kumar V, Fausto N, Robbins SL, Abbas AK Robbins and Cotran pathologic basis of disease. St. Louis, MO: Elsevier Saunders, 2005;715.
6. Peres e Serra A, Parra ER, Eher E, Capelozzi VL. Nonhomogeneous immunostaining of hyaline membranes in different manifestations of diffuse alveolar damage. Clinics 2006;61(6):497–502.
7. Ware L, Matthay M. The acute respiratory distress syndrome. N Engl J Med 2000;342(18):1334–49.
8. Abraham E, Singer M. Mechanisms of sepsis-induced organ dysfunction. Crit Care Med 2007;35(10):2408–16.
9. Levy MM, Fink MP, Marshall JC, Abraham E, Angus D, Cook D, et al. 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference. Crit Care Med 2003;31(4):1250–6.
10. Sloane PJ, Gee MH, Gottlieb JE, Albertine KH, Peters SP, Burns JR, et al. A multicenter registry of patients with acute respiratory distress syndrome. Physiology and outcome. Am Rev Respir Dis 1992;146(2):419–26.
11. Amato M, Barbas C, Medeiros D, Magaldi R, Schettino G, Lorenzi-Filho G, et al. Effect of a protective-ventilation strategy on mortality in the acute respiratory distress syndrome. N Engl J Med 1998;338(6):347–54.
12. Chapman HC. Docimasia pulmonum hydrostatica. In: Chapman HC, editor. A manual of medical jurisprudence, insanity and toxicology, 3rd edn. Philadelphia, New York, London: Saunders & Company, 1903;192–4.
13. Turillazzi E, Di Donato S, Neri M, Riezzo I, Fineschi V. An immunohistochemical study in a fatal case of acute interstitial pneumonitis (Hamman-Rich syndrome) in a 15-year-old boy presenting as sudden death. Forensic Sci Int 2007;173(1):73–7.
14. Turillazzi E, Karch SB, Neri M, Pomara C, Riezzo I, Fineschi V. Confocal laser scanning microscopy. Using new technology to answer old questions in forensic investigations. Int J Legal Med 2008;122(2):173–7.
15. Ridgway EJ. The microbiology of the autopsy. In: Burton JL, Rutty GN, editors. The hospital autopsy, 2nd edn. London: Arnold, 2001;134–46.
16. de Roux A, Marcos MA, Garcia E, Mensa J, Ewig S, Lode H, et al. Viral community-acquired pneumonia in nonimmunocompromised adults. Chest 2004;125(4):1343–51.
17. Beigel JH, Farrar J, Han AM, Hayden FG, Hyer R, de Jong MD, et al. Writing Committee of the World Health Organization (WHO) Consultation on Human Influenza A/H5. Avian influenza A (H5N1) infection in humans. N Engl J Med 2005;353(13):1374–85.
18. Bauer TT, Ewig S, Rodloff AC, Müller EE. Acute respiratory distress syndrome and pneumonia: a comprehensive review of clinical data. Clin Infect Dis 2006;43(6):748–56.
19. Batalis NI, Caplan MJ, Schandl CA. Acute deaths in nonpregnant adults due to invasive streptococcal infections. Am J Forensic Med Pathol 2007;28(1):63–8.
20. Pelosi P, Caironi P, Gattinoni L. Pulmonary and extrapulmonary forms of acute respiratory distress syndrome. Semin Respir Crit Care Med 2001;22(3):259–68.
21. Esteban A, Fernández-Segoviano P, Frutos-Vivar F, Aramburu JA, Nájera L, Ferguson ND, et al. Comparison of clinical criteria for the acute respiratory distress syndrome with autopsy findings. Ann Intern Med 2004;141(6):440–5.
22. Kobayashi H, Itoh T, Sasaki Y, Konishi J. Diagnostic imaging of idiopathic adult respiratory distress syndrome (ARDS)/diffuse alveolar damage (DAD). Hystopathologic correlation with radiological imaging. Clin Imaging 1996;20:1–7.
23. Goodman LR, Fumagalli R, Tagliabue P, Tagliabue M, Ferrario M, Gattinoni L, et al. Adult respiratory distress syndrome due to pulmonary and extrapulmonary causes: CT, clinical, and functional correlations. Radiology 1999;213(2):545–52.
24. Tomashefski JF Jr. Pulmonary pathology of acute respiratory distress syndrome. Clin Chest Med 2000;21(3):435–66.
25. Brun-Buisson C, Minelli C, Bertolini G, Brazzi L, Pimentel J, Lewandowski K, et al. Epidemiology and outcome of acute lung injury in European intensive care units. Results from the ALIVE study. Intensive Care Med 2004;30(1):51–61.
26. Tsokos M. Post-mortem diagnosis of sepsis. Forensic Sci Int 2007;165:155–64.
27. Mandell LA, Bartlett JG, Dowell SF, File TM Jr, Musher DM, Whitney C, et al. Update of practice guidelines for the management of community-acquired pneumonia in immunocompetent adults. Clin Infect Dis 2003;37(11):1405–33.
28. Estenssoro E, Dubin A, Laffaire E, Canales H, Sáenz G, Moseinco M, et al. Incidence, clinical course, and outcome in 217 patients with acute respiratory distress syndrome. Crit Care Med 2002;30(11):2450–6.

Additional information and reprint requests:
Francesco Ventura, M.D., Ph.D.
Department of Legal and Forensic Medicine
University of Genova
Via de Toni 12
16132 Genova
Italy
E-mail: francesco.ventura@unige.it

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01437.x
Available online at: interscience.wiley.com

CASE REPORT

PATHOLOGY AND BIOLOGY

Michael Panella,[1] *M.D., J.D.; Janice E. Ross,*[2] *M.D.; Keene Garvin,*[2] *M.D.; and Alex Martin,*[2] *B.S.*

Cardiac Sudden Death as a Result of Acute Coronary Artery Thrombosis During Chemotherapy for Testicular Carcinoma

ABSTRACT: We report the first acute coronary fibrin thrombus arising upon atherosclerosis detected at autopsy in a man receiving chemotherapy for testicular carcinoma. The decedent was a smoker with no other known atherosclerotic risk factors. Histology revealed superficial atherosclerotic plaque erosion with endothelial necrosis and no intraplaque hemorrhage. A focus of intimal lymphoid infiltrates was noted away from the plaque. These findings raise the possibility of chemotherapy-induced vascular damage as a factor in thrombogenesis. A review of Pubmed was performed which documented clinical reports of an association of chemotherapy with acute cardiac ischemia but no well described autopsy findings. Our case highlights the need for careful assessment of the coronary system in chemotherapy patients dying suddenly, particularly in the absence of significant atherosclerotic risk factors. Such postmortem examination will ensure thorough death investigation and may elucidate the pathogenesis of thrombosis with potential reduction in cardiac ischemic risks of chemotherapy patients.

KEYWORDS: forensic science, coronary thrombus, chemotherapy, atherosclerosis, sudden death and autopsy

Cardiac sudden death arising from atherosclerotic coronary artery disease is a very common finding at autopsy. All are familiar with the underlying mechanism for such deaths which centers on myocardial ischemia arising from a mismatch of myocardial demand and blood supply. This mismatch usually depends on a complicated interaction of coronary atherosclerosis, plaque rupture, clotting, and vasospasm. Given the importance of atherosclerosis in this interaction, many forensic pathologists strongly consider ischemic heart disease in the investigation of sudden death when risk factors linked to atherosclerosis are present, including age, hypertension, diabetes, smoking, hypercholesterolemia, and familial history. Yet, some individuals succumb from coronary artery insufficiency without the typical atherosclerotic risk factors which makes diagnosis particularly challenging. Such cases include Prinzmetal vasospasm, drugs, paradoxical emboli from the right side of the heart via a septal defect, vasculitis, amyloidosis, and hemoglobinopathies (1).

From a forensic standpoint, many pathologists readily associate drug-related ischemic heart disease with cocaine or other vasospasm-inducing medications. However, another less widely appreciated drug class associated with ischemic heart disease is oncological chemotherapeutic agents which may cause vascular toxicity. We present an autopsy case of sudden cardiac death occurring during chemotherapy for testicular carcinoma, which is secondary to an acute coronary thrombosis arising upon occult atherosclerosis. The decedent did not possess significant atherosclerotic risk factors, other than male sex and tobacco use, which would make the death suspicious for atherosclerotic coronary artery disease. The medical literature is reviewed via a Pubmed search in terms of ischemic heart disease and testicular carcinoma chemotherapy. After review of the literature, several clinical reports of coronary artery insufficiency are noted; however, our finding of an acute fibrin thrombus arising upon an atherosclerotic plaque during chemotherapy is the first reported case documented at forensic autopsy. This finding may imply that young patients with asymptomatic coronary atherosclerotic plaques are at increased risk of ischemic heart disease and sudden cardiac death during testicular carcinoma chemotherapy. Furthermore, our case highlights the importance of considering ischemic heart disease after a sudden death during testicular carcinoma chemotherapy with careful autopsy examination of the coronary system in such victims for atherosclerotic-related thrombosis, even in the absence of significant risk factors for atherosclerosis. In addition, given reports of acute myocardial ischemia in other chemotherapy-treated patients with nontesticular carcinoma, this heightened suspicion of cardiac sudden death should extend to all chemotherapeutic regimens.

Our histopathological findings of the coronary artery thrombosis arising upon an atherosclerotic plaque are also presented, which raise the possibility of vascular toxicity as a factor in thrombogenesis during chemotherapy. Additional cases detected at autopsy may confirm this possibility and further elucidate the roles of occult atherosclerosis, vascular toxicity, and thrombosis in chemotherapy-induced myocardial ischemia. Consequently, greater awareness and vigilance at autopsy for coronary thrombosis during chemotherapy is important not only in accurate death investigation but also in increasing our understanding of the pathophysiology of

[1]University of Missouri, Department of Pathology and Anatomical Sciences, Columbia, MO.
[2]Newberry Pathology Associates, 185 Executive Drive, Newberry, SC.

Received 5 June 2009, and in revised form 24 July 2009, accepted 4 Aug. 2009.

chemotherapy-induced myocardial ischemia. With this greater understanding, better pretreatment screening for cardiac risks may be possible in patients undergoing chemotherapy treatment to reduce cardiac sudden death during therapy.

Case Report

The decedent was a 37-year white male recently diagnosed with a probable Stage I B (T2, N0, Mx) nonseminomatous testicular carcinoma of his right testicle. His past medical history was significant for tobacco use consisting of a pack of cigarettes for 16 years. There was no reported history of hypercholesterolemia, hypertension, or diabetes. He underwent a right orchiectomy but was noted to have persistent elevation of his α-fetoprotein serum level. Consequently, chemotherapy was initiated using a bleomycin, etoposide, and cisplatin (BEP) protocol. The decedent received three doses of cisplatin with his last dose 3 days before his death. On the day of his death, he received a dose of blenoxane in an out-patient setting. He was sent home feeling fine and went for a walk outside of his residence. During his walk, he experienced chest pain and notified his treating physician who recommended an analgesic pill for possible chemotherapy-induced musculoskeletal pain. Later that day, the decedent collapsed and was rushed to a nearby hospital in cardiac arrest. An electrocardiogram in the emergency room showed a possible acute lateral infarction. Despite advanced life support, he was eventually pronounced dead. The decedent was referred to our office for an autopsy by the local coroner given the recent infusion of chemotherapy and the lack of significant past cardiac problems which could adequately explain the death.

Autopsy revealed a thin man with an ideal body mass index of 22.32. The heart was 400 g with no evidence of hypertensive hypertrophy or gross myocardial scarring. The coronary arteries demonstrated minimal gross coronary artery atherosclerosis of the proximal left anterior descending artery with 20% occlusion. Just distal to this gross atherosclerotic plaque, an acute thrombus was identified which completely occluded the lumen (Fig. 1). Microscopic sections of this area confirmed an acute fibrin thrombus and revealed underlying atherosclerosis (Fig. 2). This atherosclerotic plaque was not appreciated on gross examination given the obscuring fibrin thrombus. Microscopically, this atherosclerotic plaque possessed approximately 50% occlusion with no evidence of deep intraplaque hemorrhage or rupture. There was superficial plaque erosion with patchy overlying coagulative necrosis of the endothelial cells (Figs. 3 and 4). In addition, the intima possessed rare lymphocytic infiltrates within the zone of fibrin thrombus but away from the atherosclerotic plaque (Fig. 5). The myocardium was free of microscopic fibrosis and acute infarction changes. The lungs were congested with a combined weight of 1765 g. The right testicle was absent with a fibrotic spermatic cord stump that was consistent with previous surgical removal. Light microscopy of this fibrotic stump revealed degenerating residual nonseminomatous carcinoma (Fig. 6). The remainder of the autopsy was essentially unremarkable with no evidence of other occult malignancy. Postmortem toxicology was significant for Tetrahydrocannabinol and Benadryl.

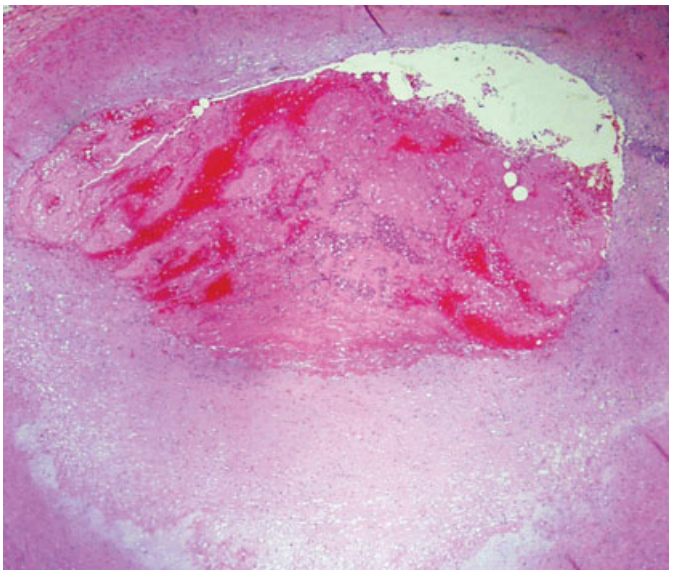


FIG. 2—*Microscopic section of acute fibrin thrombus with underlying atheromatous plaque (H&E 2X).*

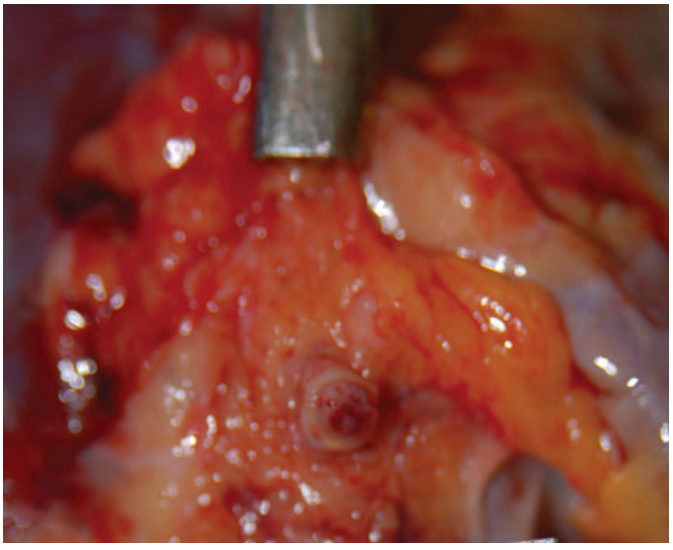


FIG. 1—*Grossly appearing fibrin thrombus of proximal left anterior descending artery.*

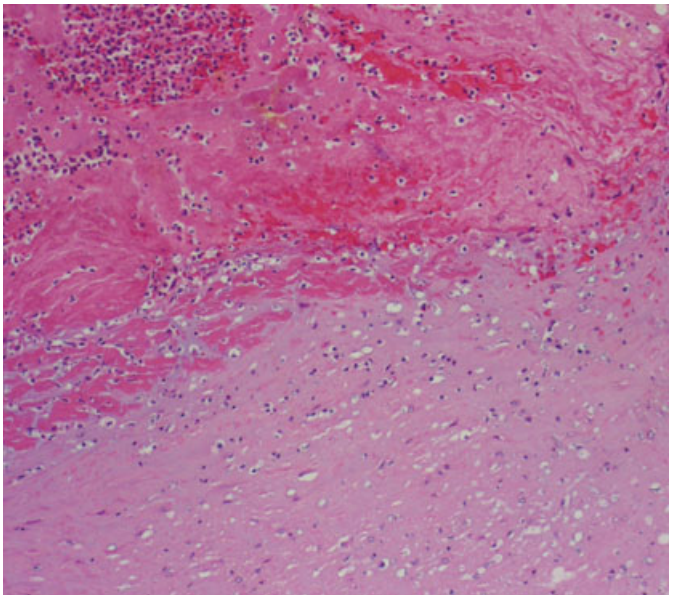


FIG. 3—*Superficial plaque erosion (H&E 10X).*

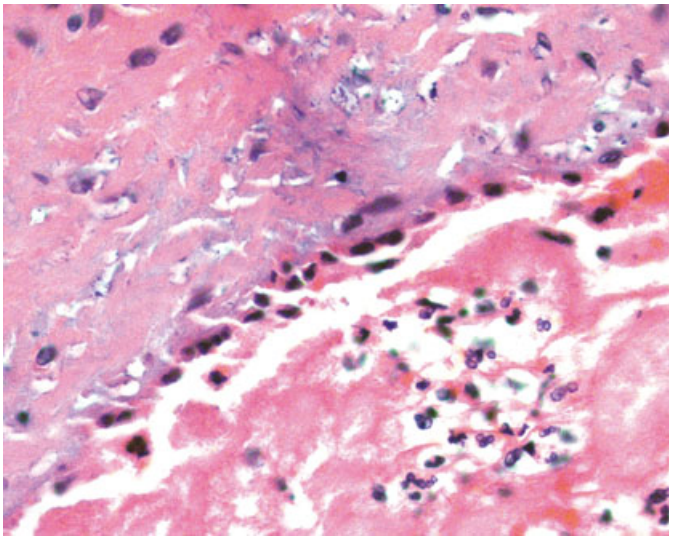


FIG. 4—*Endothelial coagulative necrosis (H&E 40X).*

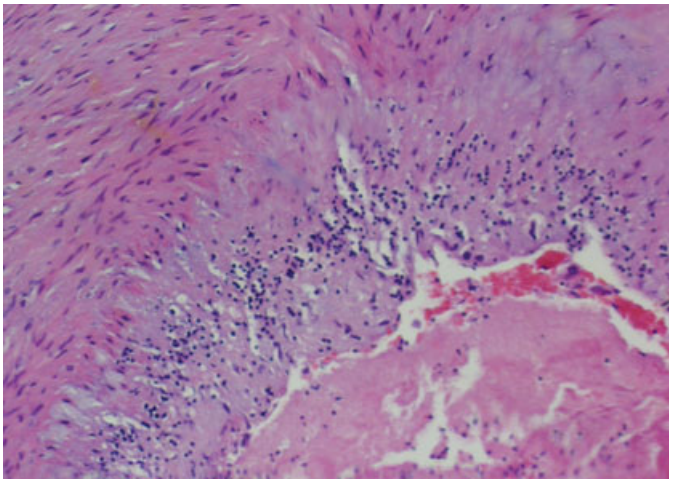


FIG. 5—*Intimal lymphoid infiltrate (H&E 20X).*

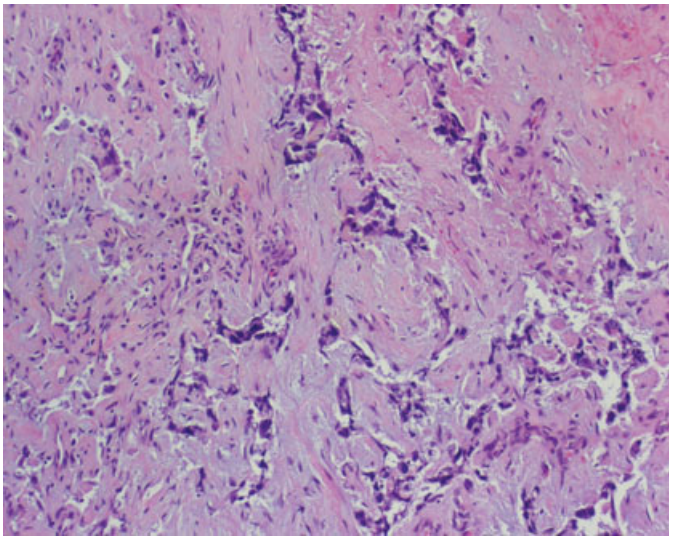


FIG. 6—*Residual partially degenerating non-seminomatous carcinoma of right spermatic cord stump.*

Comment

At first glance, this case seems like a typical cardiac sudden death arising from acute fibrin thrombus because of atherosclerotic coronary artery disease. However, further examination of these autopsy findings raises the question of whether the acute fibrin thrombus is related to the ongoing chemotherapy. A review of Pubmed literature for myocardial ischemia occurring during testicular carcinoma chemotherapy demonstrated that such an association exists with several clinical case reports documenting acute cardiac events during such therapy. However, no cardiac sudden deaths or a description of the histopathological coronary findings were reported on such patients with cancer at the time of forensic autopsy. Consequently, our case represents the first reported forensic autopsy of a sudden cardiac sudden death occurring during testicular carcinoma treatment with histopathological findings of an acute coronary thrombus arising on an atherosclerotic plaque.

The majority of the previous Pubmed articles report myocardial infarctions diagnosed by clinical modalities. Besides these documented myocardial infarction cases, a single clinical report also noted various noninfarct vascular ischemia associated with bleomycin, vinblastine, and cisplatin testicular carcinoma chemotherapy, which included angina pectoris, Raynaud's phenomenon, toe ischemia, and migraines (2). The reports of myocardial infarctions associated with testicular carcinoma chemotherapy begin in late 1980s and involve at least 16 patients. Four of these individuals received bleomycin, etoposide, and cisplatin with one of these patients also simultaneously suffering two acute cerebral infarcts (3–6). Another clinical report noted a non-Q wave myocardial infarction associated with bleomycin and etoposide (7). Infarct-like changes consisting of elevated heart enzymes and apical dyskinesia on ventriculogram but normal coronary arteries on angiogram were documented in a patient with embryonic carcinoma of the testes who received vinblastine, bleomycin, and cisplatin (8). In one case, angiography with atherectomy confirmed a thrombus (6). However, none of these clinical reports described histopathological findings of an acute thrombus arising within a preexisting atherosclerotic plaque.

Although the acute fibrin thrombus detected in our case may be a natural progression of the underlying atherosclerosis, it is difficult to exclude a causal relationship of the thrombosis with the ongoing chemotherapy. This causal relationship is supported by the temporality of the chemotherapy treatment to the death and the relative absence of multiple significant atherosclerotic risk factors in our case. Indeed, the previous clinical reports support an association of the myocardial ischemia with testicular carcinoma chemotherapy possibly related to vascular damage from the chemical agents. In fact, vascular toxicity associated with chemotherapy has been known for at least three decades. In particular, acute arterial ischemic events were noted to occur most frequently after cisplatin-based combination regimens with possible mechanisms including drug-induced endovascular damage, perturbation of the clotting system, platelet activation, thromboxane-prostracyclin homeostasis abnormalities, autonomic dysfunction, vasculitis, and stimulation of fibroblasts (9).

Based on two histopathology findings, our autopsy tends to support vascular damage as a potential factor in thrombogenesis arising in atherosclerotic plaques during testicular carcinoma chemotherapy. First, the atherosclerotic coronary plaque possesses superficial erosion with superimposed acute fibrin thrombi, which raises the possibility of superficial plaque disruption with subsequent fibrin deposition and thrombus formation. Further support for superficial plaque erosion as the source for the thrombosis is based on the absence of well-defined deep intraplaque hemorrhage or rupture,

which is classically associated with mechanical shearing forces as the etiology for acute coronary fibrin thrombus. This surface disruption may be secondary to the endothelial coagulative necrosis that was documented within the areas of the atherosclerotic plaque. Such endothelial necrosis may be a product of the toxic effects of the chemotherapy. A similar mechanism of endothelial damage is postulated to be a source of fluoruracil-induced erosions and ulcerations of the ileum in which thrombogenic and vasospastic effects of the chemotherapeutic agent occur on the vascular epithelium (10). Second, further evidence of potential chemotherapy-induced vascular damage centers on the intimal lymphocytic infiltrate identified in our case, which was noted within the zone of fibrin thrombus but away from the plaque. Although intimal inflammation may be a secondary process in the development and progression of an atherosclerosis, one would expect to find such secondary inflammation beneath and immediately along the sides of the atherosclerotic plaque. Our histology demonstrated the intimal lymphoid infiltrate away from the plaque that argues for a primary inflammatory vasculitic response possibly related to the chemotherapy. Vasculitis associated with chemotherapy has been potentially linked with a couple of agents, such as gemcitabine in pancreatic cancer where a possible early skin leukocytoclastic vasculitis has been described and bleomycin in which skin rash biopsies have demonstrated a spectrum of findings including lymphocytic vasculitis (11,12). Given that our decedent had a course of bleomycin prior to death, the argument for a primary vasculitis as the source of the intimal lymphoid infiltrate is strengthened.

This case highlights the need for pathologists to consider acute coronary fibrin thrombi in sudden death victims who received testicular cancer chemotherapy, particularly in the absence of significant atherosclerotic risk factors. Without an awareness of a possible association of chemotherapy with acute fibrin thrombi arising in atherosclerotic plaque, pathologists may miss acute cardiac ischemic deaths arising from chemotherapy. Such misdiagnoses are problematic not only for pathologists but also for clinicians as demonstrated by our case in which the decedent called his oncologist for chest pain and was told to take an analgesic without evaluation for possible cardiac ischemia. Given the importance of death investigation in promoting a better understanding of treatment risks and preventing future misdiagnoses, pathologists should carefully consider the possibility of acute thrombi arising in occult coronary atherosclerotic plaques during chemotherapy and document such findings.

This documentation of the coronary system in chemotherapy patients may permit further elucidation of the pathogenesis of cardiac-related deaths associated with testicular carcinoma treatment. Such understanding may permit better cardiovascular risk assessment of patients prior to chemotherapy to potentially reduce the occurrence of acute ischemic events during treatment. After all, if underlying coronary atherosclerosis with superimposed vascular toxicity is important in the pathogenesis of acute thrombosis during chemotherapy, our decedent, whose only major risk factor was tobacco use, may serve as example in which tobacco exposure alone increases the probability of chemotherapy-associated acute myocardial ischemia. Such a history of tobacco use may require additional screening prior to and during chemotherapy for the detection and prevent of acute myocardial ischemia. Consequently, all sudden death victims with a history of chemotherapy treatment should have a careful dissection of the coronary arteries at autopsy to detect and evaluate any pathological lesions producing myocardial ischemia, including vascular damage, fibrin thrombi, and atherosclerosis.

It must be noted that this postmortem coronary evaluation is important not just with testicular carcinoma chemotherapy regimens but should include all chemotherapy-treated patients given several factors. First, cisplatin-based chemotherapy associated with acute myocardial infarction is not just linked to testicular carcinoma but includes treatment for malignant melanoma (13). Second, other chemotherapy-treated patients with nontesticular carcinoma may possess an increased risk of ischemic heart-related complications. Acute myocardial infarctions have been described with 5-fluorouracil therapy for pancreatic carcinoma and following paclitaxel therapy for ovarian carcinoma (14,15). A recent article also notes that the newer chemotherapy agent gemcitabine is associated with important thrombotic and vascular side effects that may be higher than originally thought (16). Finally, the risk of premature myocardial infarctions is not only centered within the immediate period of treatment but has been reported to extend several months to years beyond the therapy (17). Consequently, careful postmortem evaluation of the coronary system may be prudent in all chemotherapy patients regardless of the time interval from treatment to death.

Conclusion

We presented a case of cardiac sudden death in a young man associated with chemotherapy for testicular carcinoma in which his only risk cardiac factor was male sex and tobacco use. Autopsy demonstrated an acute coronary thrombus superimposed upon an atherosclerotic plaque. Review of the medical literature showed several clinical case reports of myocardial infarction and noninfarct vascular ischemia associated with testicular carcinoma chemotherapy but no reported histopathological findings documented at forensic autopsy. Our histopathology showed an atherosclerotic coronary plaque with overlying acute fibrin thrombus, superficial plaque erosion, and endothelial necrosis. In addition, rare intimal lymphocytic infiltrates were present within the zone of the thrombus but away from the atherosclerotic plaque. Deep intraplaque hemorrhage and rupture were not identified. These findings raise the possibility of chemotherapy-induced vascular damage as a factor in the pathogenesis of the acute coronary thrombus arising within an atherosclerotic background. This case will heighten the suspicions of pathologists in excluding cardiac sudden deaths during chemotherapy because of acute thrombi arising in atherosclerotic plaques, particularly in the absence of significant atherosclerotic risk factors. Careful postmortem examination of the coronary system may further elucidate the pathogenesis and promote better assessment of cardiac risks of patients undergoing chemotherapy. This postmortem coronary evaluation should include not only chemotherapy-treated patients with testicular carcinoma but all chemotherapy patients given the reports of acute myocardial infarctions for other nontesticular carcinoma chemotherapy drugs. As a final note, given the reports of remote cardiovascular complications following chemotherapy, both current and past chemotherapy patients should have postmortem evaluation of the coronary system for sudden death investigations.

Acknowledgments

We thank Ms. Sheila Martin and Ms. Jane Moose of Newberry County Memorial Hospital for their histological preparation of the microscopic sections.

References

1. Schoen FJ. The heart. In: Kumar V, Abbas A, Fausto N, editors. Robbins and Cotran pathological basis of disease, 7th edn. Philadelphia, PA: Elsevier Saunders, 2005;576.

2. Stefenelli T, Kuzmitis R, Ulrich W, Glogar D. Acute vascular toxicity after chemotherapy with cisplatin, vinblastine and bleomycin for testicular cancer. Eur Heart J 1988;9:552–6.
3. Bachmeyer C, Joly H, Jorest R. Early myocardial infarction during chemotherapy for testicular carcinoma. Tumori 2000;86:428–30.
4. Mermershatain W, Dudnik J, Gusakova I, Ariad S. Acute myocardial infarction in a young man receiving chemotherapy for testicular cancer: case report. J Chemother 2001;13:658–60.
5. Brouha ME, Bioemendal HJ, Kappelle LJ, Winter JB. Cerebral infarction and myocardial infarction due to cisplatin-containing chemotherapy. Ned Tijdchr Geneeskd 2003;147:457–60.
6. Shoji S, Miyakita H, Shima M, Usui Y, Nagata Y, Uchida T, et al. Acute myocardial infarction during combined chemotherapy with bleomycin, etoposide and cisplatin for testicular cancer. Hinyokika Kiyo 2006;52:723–6.
7. Schwarzer S, Eber B, Greinix H, Lind P. Non-Q wave myocardial infarction associated with bleomycin and etoposide chemotherapy. Eur Heart J 1991;12:748–50.
8. Zeymer U, Neuhaus KL. Infarct-typical changes in the electrocardiogram following chemotherapy with vinblastine. Dtsch Med Wochenschr 1989;114:589–92.
9. Doll DC, Ringenberg QS, Yarbro JW. Vascular toxicity associated with anti-neoplastic agents. J Clin Oncol 1986;4:1405–17.
10. Bucaloiu I, Dubagunta S, Pachipala K, Kamal N, Fata F. Small-cell cancers and an unusual reaction to chemotherapy, case 4 fluorouracil-related small bowel vasculitis. J Clin Oncol 2003;21:2442–3.
11. Li J, Ko Cj, Saif MW. Recurrent cutaneous toxic erythema Induced by gemcitabine in a patient with pancreatic cancer. Cutan Ocul Toxicol 2009;28:144–8.
12. Chen Y, Rahemtullah A, Breeden E, Hochberg E. Diagnosis in oncology, bleomycin-induced flagellate erythema. J Clin Oncol 2007;25:898–900.
13. Fukunga T, Soejima H, Sugamura K. Acute myocardial infarction Induced by cis-platin-based combination chemotherapy for malignant melanoma: a case report. J Cardiol 2006;47:191–5.
14. Braumann D, Mainzer K, Gunzl C, Lewerenz B. Myocardial infarcts within the scope of 5-fluorouracil therapy. Onkologie 1990;13:465–7.
15. Laher S, Karp SJ. Acute myocardial infarction following paclitaxel administration for ovarian carcinoma. Clin Oncol (R Coll Radiol) 1997;9:124–6.
16. Dasanu CA. Gemcitabine:vascular toxicity and prothrombotic potential. Expert Opin Drug Saf 2008;7:703–16.
17. Borek G, Gebel F, Jenzer HR. Acute myocardial infarct following chemotherapy for testicular carcinoma: on the coronary toxicity of cytostatic agents. Schweiz Med Wochenschr 1991;121:385–9.

Additional information and reprint requests:
Michael Panella, M.D., J.D.
University of Missouri Department of Pathology and
Anatomical Sciences
1 Hospital Drive
M263 Medical Science Building
Columbia MO 65212
E-mail: panellam@health.missouri.edu

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01410.x
Available online at: interscience.wiley.com

CASE REPORT

PATHOLOGY AND BIOLOGY

Marianna Shvartsbeyn,[1] *M.D.; Daniel G. K. Phillips,*[2] *M.D.; Michael A. Markey,*[3] *M.D.; Alan Morrison,*[2] *M.D.; Joyce L. DeJong,*[3] *M.D.; and Rudy J. Castellani,*[2] *M.D.*

Cocaine-Induced Intracerebral Hemorrhage in a Patient with Cerebral Amyloid Angiopathy*

ABSTRACT: Intracerebral hemorrhage (ICH) is a well-recognized complication of recreational cocaine use. The precise mechanism of the cocaine-induced hemorrhagic event is unclear, although multiple factors have been implicated. We report a case of a 62-year-old woman who suffered left parieto-occipital ICH with herniation and death, following a cocaine binge. Microscopic examination also revealed extensive cerebral amyloid angiopathy (CAA) in the vicinity of the hemorrhage. We additionally studied brain tissue in eight subjects between ages of 60 and 80 who were positive for cocaine metabolites at autopsy; of these, none had vascular amyloid-β deposits by immunohistochemistry. Whereas we found no evidence that chronic cocaine use is a risk factor for CAA, given the age-associated nature of CAA and the aging population using cocaine, CAA-induced hemorrhage in the setting of cocaine use may be more common than recognized. This is the first reported case of CAA-associated ICH precipitated by cocaine.

KEYWORDS: forensic science, cocaine, intracerebral hemorrhage, amyloid beta, immunohistochemistry, cerebral amyloid angiopathy, neurovascular complications, review

Intracerebral hemorrhage (ICH) is a well-recognized complication of cocaine use. Cocaine remains one of the leading causes of drug-related deaths (1) and drug-related visits to the emergency department (2). Because of its sympathomimetic effects, it causes arterial vasospasm and secondary thrombus leading to cerebral infarction (3) and/or a transient spike in blood pressure predisposing to intracranial hemorrhage (4). Approximately half of cocaine users who suffered ICH had an underlying cerebral saccular aneurysm or arteriovenous malformation (AVM) (5). Vasculitis was suggested as another mechanism of cocaine-induced hemorrhagic stroke based on angiographic findings of "beading" in multiple cerebral arteries which was proportional to the drug exposure, although pathological vasculitis secondary to acute cocaine toxicity has never been convincingly demonstrated (6–8).

Cerebral amyloid angiopathy (CAA), in contrast, is the leading cause of lobar ICH in nonhypertensive subjects over the age of 60 (1,9–12). Even mild amyloid deposits apparently cause significant vessel dilation and alter the cerebral microvasculature (13). Nearly 20% of patients with CAA suffer hemorrhage, while subclinical microhemorrhages on T2-weighted MRI are found in up to 80% of people with CAA (14). It is hypothesized that cerebral hemorrhage associated with CAA is initiated by seeding and extracellular deposition of Aβ40 (15).

The population of individuals using cocaine is aging (1), and CAA remains a common disease process associated with advanced age. The relationship between the two has never been explored nor has the association between cocaine and CAA-associated hemorrhage been reported.

Case Report

The decedent was a 61-year-old white woman with a history of bipolar disorder, asthma, chronic obstructive pulmonary disease, panic attacks, ethanol, and recreational cocaine use who was found dead at home. Postmortem urine toxicology revealed a presence of benzoylecgonine in the urine. The patient had a witnessed cocaine binge, and she became unresponsive during her binge. She died immediately, which would not have allowed enough time for the cocaine to metabolize; along with the presence of benzoylecgonine in the urine, and the fact that she was not hypertensive (heart weight 320 g, kidneys without hypertensive changes, no history of hypertension) this is a strong enough evidence to ascribe the bleed to the cocaine. The most remarkable gross finding at autopsy was the presence of a large left parieto-occipital (lobar) ICH (Fig. 1). There was a significant mass effect, including left transtentorial, left to right subfalcine, and cerebellar tonsillar herniation. Microscopic examination confirmed the acute hemorrhage and, in addition, demonstrated extensive replacement of cortical vessels by amorphous eosinophilic material (Fig. 2*A*). The material was Congo red positive with red-green birefringence and was strongly immunoreactive using antibodies against residues 17–24 of the amyloid-β peptide (Fig. 2*B*), indicative of CAA. The postmortem urine drug screen had evidence of cocaine use. The general autopsy was otherwise unremarkable.

The finding of CAA in an older cocaine user raised the question of whether chronic cocaine use was a risk factor for CAA. In an attempt to answer this question, we searched our records for other subjects in whom cocaine metabolites were found at autopsy, and who were also over the age of 60 with brain tissue available for

[1]Department of Pathology, New York University School of Medicine, New York, NY.
[2]Division of Neuropathology, University of Maryland, Baltimore, MD.
[3]Division of Forensic Pathology, Sparrow Health System, Lansing, MI.
*Presented at the 60th Annual Meeting of the American Academy of Forensic Sciences, February 18–23, in Washington, DC.

Received 10 Mar. 2009; and in revised form 28 June 2009; accepted 10 July 2009.

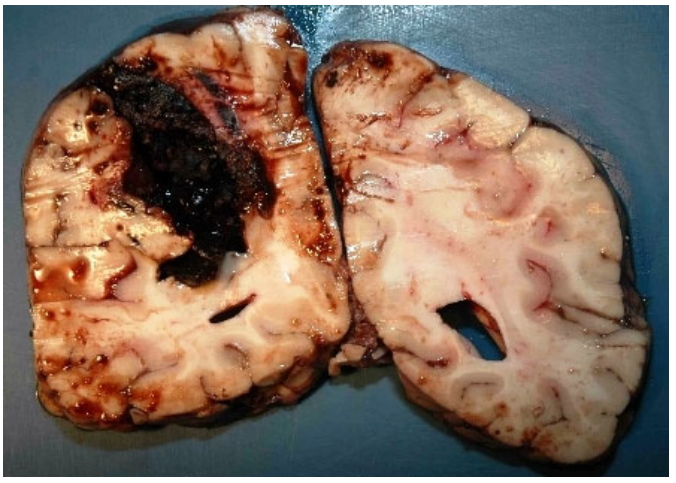


FIG. 1—*Coronal section of the cerebral hemispheres posterior to the splenium showing a large lobar intracerebral hematoma in the left parietal region, with mass effect.*

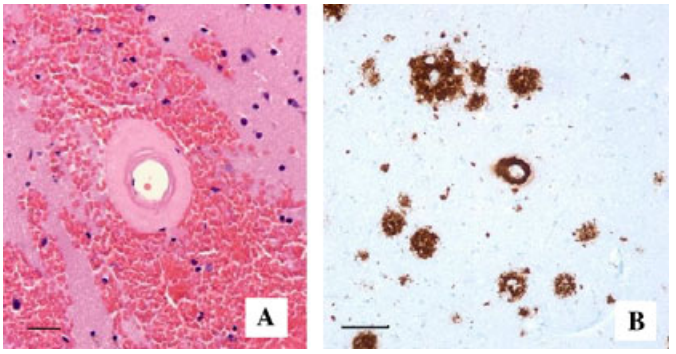


FIG. 2—*Microscopic sections demonstrate extensive amyloid within cerebral vessels (A; scale bar = 20 μm), confirmed by amyloid-β immunohistochemistry (B; scale bar = 50 μm).*

examination. Eight subjects were found. The cerebral cortex of each decedent was examined and immunostained for amyloid-β. No significant amyloid-β accumulations were found in the cerebral vasculature in any of the eight subjects, while only one 80-year-old subject had scattered diffuse plaques (an age-related finding) with no CAA. This study was presented to our Institutional Review Board and exempted from further review.

Discussion

Intracerebral hemorrhage is a well-recognized complication of recreational cocaine use. The hemorrhage ensues within minutes to hours after exposure to the drug (12). While such subjects often have underlying structural lesions (saccular aneurysms, AVMs), a significant number of cocaine-associated strokes are "spontaneous" or occur in the absence of any other vascular pathology. Interestingly, different forms of cocaine are reported to be associated with different end-organ lesions: alkaloidal or "crack" cocaine was found to precipitate both hemorrhagic and ischemic events with equal frequencies, whereas cocaine hydrochloride caused ICH four times more frequently than vascular ischemia (5).

Subarachnoid hemorrhage (SAH) and intraparenchymal hemorrhage (IPH) are the most commonly described forms of ICH associated with cocaine use (16). In a recent review of combined SAH and IPH related to cocaine abuse, no underlying anatomic abnormality was identified in about half of the cases (17). However, the incidence of ruptured intracranial aneurysms in cocaine users who suffer SAH has been estimated at 85% (18). In one series, a saccular aneurysm was found in 78% of cocaine-related SAH, most of which involved anterior or posterior communicating arteries (19). In another series, 12 patients with cocaine-induced SAH all had an underlying cerebral aneurysm (18). Similar analysis performed on 16 subjects who presented with neurovascular complications of cocaine abuse, revealed SAH and IPH in four patients each (16). In the SAH group, only one patient did not have an aneurysm diagnosed by either angiography or autopsy. One of the patients with cocaine-induced IPH had a vascular malformation. Per a literature review presented by Jacobs et al. (16), seven other cases of cocaine-related SAH were all associated with a finding of berry or saccular aneurysms. Few cases of IPH were associated with an underlying AVM or a brain neoplasm (16).

The anatomical distribution of the cocaine-induced intracerebral bleeding varies. In one single-institution analysis, nine patients had subarachnoid, 16 intracerebral (eight basal ganglia, seven hemispheric, and one brain stem), and five intraventricular hemorrhages (IVHs). Six patients had aneurysms, and three had AVMs (20). Two other cases of pontine hemorrhage secondary to cocaine use were described (21,22).

When all autopsies of nontraumatic intracranial hemorrhage were analyzed in one institution over a period of 1 year, 10 of the 17 cases identified had evidence of cocaine use antemortem (23). Seven suffered IPH, while the rest had SAH, all associated with saccular aneurysms. Of 33 patients who presented to the Detroit Medical Center with neurological symptoms in a setting of cocaine use, 15 suffered hemorrhagic neurovascular events. Five of eight patients with SAH were found to have aneurysms, predominantly involving the posterior communicating artery. Four patients had IPH extending to the ventricles of the subarachnoid space, and three patients had an isolated IPH. Two patients had AVMs, one in the temporoparietal location and one in the conus of the spinal cord (24). In another study, 12 patients who underwent cerebral arteriography following SAH were found to have underlying intracerebral aneurysms (25). Six patients had evidence of ICH in parietal, frontotemporal, pontine, or basal ganglia areas; three of them underwent cerebral arteriography that revealed an AVM in one of them. One patient suffered an IVH.

A high percentage of intracerebral hematomas involving the basal ganglia and thalamus are noted in IPH temporally related to cocaine use (16,19,26). Extensive deep hemorrhages into the basal ganglia or corona radiata may also be associated with IVH (26). Isolated IVH has rarely been reported in adults abusing cocaine (20,25). However, it is frequently seen in cocaine-exposed neonates (27,28).

Cerebral vasospasm induced by cocaine is likely a predisposing factor for aneurysmal rupture (18,29–32). The size of ruptured aneurysms associated with cocaine use tends to be smaller than the size of ruptured aneurysms not associated with cocaine use (18). One retrospective study looked at 108 patients with SAH who had a known aneurysm. Thirty-six of them had recent cocaine use prior to ICH, and this group had significantly higher blood pressures on admission, greater risk of developing vasospasm, and worse overall outcome (33). Similar results were reported elsewhere (34). However, the correlation is not always straightforward. No predisposing anatomical abnormalities were found in 12 autopsy cases of cocaine-induced SAH and IPH (35). No vascular changes were found in two fatal incidences of methamphetamine ingestion leading to extensive bilateral SAH in one case and a combination of subarachnoid, intraventricular, and ICH in another case (36,37).

In this report, we present the case of a 60-year-old woman who suffered a large lobar ICH following a cocaine binge and also was found to have extensive CAA. Surprisingly, this has not been previously reported in the literature; surprising because CAA is a common vascular pathology in individuals over the age of 60, is the most common cause of lobar ICH in the elderly, and cocaine use is becoming more common in older individuals. According to the 2007 report from the Substance Abuse and Mental Health Services Administration, there is a trend toward increasing cocaine use with older age (38). The highest prevalence for cocaine use was noted for individuals born between the late 1940s and early 1960s, i.e., the ''baby boomers'' (39). As this population continues to age and use cocaine, increased education of this association and possible poor outcome may be warranted.

On the other hand, a survey of eight additional cocaine users over the age of 60 demonstrated no incidental CAA in brain tissue. Therefore, while the numbers are small and preclude definitive interpretation, we cannot conclude that cocaine use per se accelerates or otherwise represents a risk for CAA apart from advanced age. Nor we can conclude that cocaine use is ''protective'' against CAA. It is nevertheless reasonable to speculate that CAA, once present, predisposes to vascular rupture and ICH in individuals who use cocaine.

In conclusion, we present the case of a 60-year-old woman who died of ICH following a cocaine binge and was found to have extensive CAA at autopsy. Whereas the association between cocaine-induced ICH and specific underlying structural lesions is well documented (e.g., saccular aneurysms, AVMs), this report represents the first case reported in the literature with cocaine-precipitated ICH in a patient with underlying CAA. Given the novel nature of this phenomenon, it will be important to carefully document the presence of cocaine in the future reports utilizing confirmatory GC/MS and blood cocaine levels. We also recommend calculating a pharmacovigilance score to quantitate the relationship between the cocaine use and the pathology found. An association between cocaine use and ICH in patients with underlying CAA may increase in frequency, given the aging population, and the aging population using cocaine. Moreover, we suspect that this association is currently under recognized and increased vigilance is warranted so that more cases can be recognized, and the aging population properly educated.

References

1. Substance Abuse and Mental Health Services Administration. Drug Abuse Warning Network, 2003: area profiles of drug-related mortality. DAWN Series D-27, DHHS Publication No. (SMA) 05-4023, Rockville, MD, March 2005.
2. Substance Abuse and Mental Health Services Administration. Drug Abuse Warning Network, 2005: national estimates of drug-related emergency department visits. DAWN Series D-29, DHHS Publication No. (SMA) 07-4256, Rockville, MD, March 2007.
3. Konzen JP, Levine SR, Garcia JH. Vasospasm and thrombus formation as possible mechanisms of stroke related to alkaloidal cocaine. Stroke 1995;26(6):1114–8.
4. Bruno A. Cerebrovascular complications of alcohol and sympathomimetic drug abuse. Curr Neurol Neurosci Rep 2003;3(1):40–5.
5. Levine SR, Brust JC, Futrell N, Brass LM, Blake D, Fayad P, et al. A comparative study of the cerebrovascular complications of cocaine: alkaloidal versus hydrochloride—a review. Neurology 1991;41(8):1173–7.
6. O'Connor AD, Rusyniak DE, Bruno A. Cerebrovascular and cardiovascular complications of alcohol and sympathomimetic drug abuse. Med Clin North Am 2005;89(6):1343–58.
7. Kaye BR, Fainstat M. Cerebral vasculitis associated with cocaine abuse. JAMA 1987;258(15):2104–6.
8. Morrow PL, McQuillen JB. Cerebral vasculitis associated with cocaine abuse. J Forensic Sci 1993;38(3):732–8.
9. Lashley T, Plant G, Rostagno A, Frangione B, Holton JL. Cerebral amyloid angiopathies: a pathologic, biochemical, and genetic view. J Neuropathol Exp Neurol 2003;62(9):885–98.
10. Revesz T, Holton JL, Lashley T, Plant G, Rostagno A, Ghiso J, et al. Sporadic and familial cerebral amyloid angiopathies. Brain Pathol 2002;12(3):343–57.
11. Okazaki H, Reagan TJ, Campbell RJ. Clinicopathologic studies of primary cerebral amyloid angiopathy. Mayo Clin Proc 1979;54(1):22–31.
12. Ferrer I, Kaste M, Kalimo H. Consequences of cerebrovascular disorders. In: Love S, Louis DN, Ellison DW, editors. Greenfield's neuropathology, 8th edn. London: Oxford University Press, 2008;206–7.
13. Kimchi EY, Kajdasz S, Bacskai BJ, Hyman BT. Analysis of cerebral amyloid angiopathy in a transgenic mouse model of Alzheimer disease using in vivo multiphoton microscopy. J Neuropathol Exp Neurol 2001;60(3):274–9.
14. Roob G, Lechner A, Schmidt R, Flooh E, Hartung HP, Fazekas F. Frequency and location of microbleeds in patients with primary intracerebral hemorrhage. Stroke 2000;31(11):2665–9.
15. Weller RO, Massey A, Newman TA, Hutchings M, Kuo YM, Roher A. Cerebral amyloid angiopathy: amyloid beta accumulates in putative interstitial fluid drainage pathways in Alzheimer's disease. Am J Pathol 1998;153(3):725–33.
16. Jacobs IG, Roszler MH, Kelly JK, Klein MA, Kling GA. Cocaine abuse: neurovascular complications. Radiology 1989;170(1 Pt 1):223–7.
17. Büttner A, Mall G, Penning R, Sachs H, Weis S. The neuropathology of cocaine abuse. Leg Med (Tokyo) 2003;5(Suppl. 1):S240–2.
18. Oyesiku NM, Colohan AR, Barrow DL, Reisner A. Cocaine-induced aneurysmal rupture: an emergent negative factor in the natural history of intracranial aneurysms? Neurosurgery 1993;32(4):518–25; discussion 525–6.
19. Green RM, Kelly KM, Gabrielsen T, Levine SR, Vanderzant C. Multiple intracerebral hemorrhages after smoking ''crack'' cocaine. Stroke 1990;21(6):957–62.
20. Daras M, Tuchman AJ, Koppel BS, Samkoff LM, Weitzner I, Marc J. Neurovascular complications of cocaine. Acta Neurol Scand 1994;90(2):124–9.
21. Ramadan NM, Levine SR, Welch KM. Pontine hemorrhage following ''crack'' cocaine use. Neurology 1991;41(6):946–7.
22. Egido-Herrero JA, Gonzalez JL. Pontine hemorrhage after abuse of cocaine. Letter. Rev Neurol 1997;25(137):137–8.
23. Nolte KB, Brass LM, Fletterick CF. Intracranial hemorrhage associated with cocaine abuse: a prospective autopsy study. Neurology 1996;46(5):1291–6.
24. Peterson PL, Roszler M, Jacobs I, Wilner HI. Neurovascular complications of cocaine abuse. J Neuropsychiatry Clin Neurosci 1991;3(2):143–9.
25. Fessler RD, Esshaki CM, Stankewitz RC, Johnson RR, Diaz FG. The neurovascular complications of cocaine. Surg Neurol 1997;47(4):339–45.
26. Nalls G, Disher A, Daryabagi J, Zant Z, Eisenman J. Subcortical cerebral hemorrhages associated with cocaine abuse: CT and MR findings. J Comput Assist Tomogr 1989;13(1):1–5.
27. Iriye BK, Asrat T, Adashek JA, Carr MH. Intraventricular haemorrhage and maternal brain death associated with antepartum cocaine abuse. Br J Obstet Gynaecol 1995;102(1):68–9.
28. McLenan DA, Ajayi OA, Rydman RJ, Pildes RS. Evaluation of the relationship between cocaine and intraventricular hemorrhage. J Natl Med Assoc 1994;86(4):281–7.
29. Howington JU, Kutz SC, Wilding GE, Awasthi D Cocaine use as a predictor of outcome in aneurysmal subarachnoid hemorrhage. J Neurosurg 2003;99(2):271–5. Comment in: J Neurosurg 2005;102(5):961-2. J Neurosurg 2008 Jun;108(6):1256-62; author reply 1262-4.
30. Vega C, Kwoon JV, Lavine SD Intracranial aneurysms: current evidence and clinical practice. Am Fam Physician 2002;66(4):601–8. Comment in: Am Fam Physician 2003 Apr 1;67(7):1438-9.
31. Vannemreddy P, Caldito G, Willis B, Nanda A Influence of cocaine on ruptured intracranial aneurysms: a case control study of poor prognostic indicators. J Neurosurg 2008;108(3):470–6. Comment in: J Neurosurg 2008 Oct;109(4):779; author reply 779-80.
32. Nanda A, Vannemreddy PS, Polin RS, Willis BK. Intracranial aneurysms and cocaine abuse: analysis of prognostic indicators. Neurosurgery 2000;46(5):1063–7. discussion 1067–9.
33. Simpson RK Jr, Fischer DK, Narayan RK, Cech DA, Robertson CS. Intravenous cocaine abuse and subarachnoid haemorrhage: effect on outcome. Br J Neurosurg 1990;4(1):27–30.
34. Aggarwal SK, Williams V, Levine SR, Cassin BJ, Garcia JH. Cocaine-associated intracranial hemorrhage: absence of vasculitis in 14 cases. Neurology 1996;46(6):1741–3.

35. McGee SM, McGee DN, McGee MB. Spontaneous intracerebral hemorrhage related to methamphetamine abuse: autopsy findings and clinical correlation. Am J Forensic Med Pathol 2004;25(4):334–7.
36. Shibata S, Mori K, Sekine I, Suyama H. Subarachnoid and intracerebral hemorrhage associated with necrotizing angiitis due to methamphetamine abuse—an autopsy case. Neurol Med Chir (Tokyo) 1991;31(1): 49–52.
37. Treadwell SD, Robinson TG. Cocaine use and stroke. Postgrad Med J 2007;83(980):389–94.
38. Substance Abuse and Mental Health Services Administration. The DASIS report: cocaine route of administration trends: 1995-2005. Rockville, MD: Substance Abuse and Mental Health Services Administration, 2007.
39. Armstrong GL. Injection drug users in the United States, 1979–2002: an aging population. Arch Intern Med 2007;167(2):166–73.

Additional information and reprint requests:
Rudy Castellani, M.D.
Department of Pathology
22 South Greene Street
Baltimore, MD 21201
E-mail: rcastellani@som.umaryland.edu

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01414.x
Available online at: interscience.wiley.com

CASE REPORT

PSYCHIATRY & BEHAVIORAL SCIENCES

J. Reid Meloy,[1,2,3] *Ph.D.*

A Catathymic Infanticide

ABSTRACT: A case of infanticide committed by a 37-year-old married man, the father of three sons, is reported. Clinically depressed since adolescence, and also diagnosed with obsessive-compulsive disorder and a dependent personality, the subject began to worry about killing someone a decade before the homicide. Increasingly disabled by his major depression, unable to work, and confined in his home, the idea that his only recourse was to kill one of his sons became fixed and frequent. Following his fourth psychiatric hospitalization, he took his 13-month-old son home from day care and drowned him in the bathtub. He then called the police and reported his crime. This sudden act of intentional killing was followed by a period of emotional relief and calmness, clearly illustrating the three stages of chronic catathymic homicide.

KEYWORDS: forensic science, forensic psychiatry, forensic psychology, catathymic process, homicide, infanticide

The intentional killing of a young child evokes disgust and horror in most, but it is even more deeply disturbing when it is inexplicable. The usual biopsychosocial conditions that can facilitate homicidal behavior—situational factors, psychosis, a toxic state, or rage—are absent, yet the intent is evident. In some such cases, the cause of the homicide is found in the faint recesses of the individual's mind, and if one looks closely, the idea of homicide has emerged slowly, almost imperceptibly, over the course of days, weeks, months, or even years, a tension filled by-product of emotion and perception that is transference-based, yet remains out of conscious awareness of the individual.

This particular motivation for homicide was originally identified by Wertham (1), who relied on earlier clinical work by Maier (2). It was called catathymia, from the Greek *kata* and *thymos*, and is most readily translated, "in accordance with emotion" (3). In a forensic context, the term refers to a motivational pattern for homicide wherein a fixed idea, often rather obsessional, grows in intensity over the course of time until the person feels compelled to kill to alleviate such psychic tension. In its chronic form, there are three clearly identifiable stages: an incubation period during which the idea, initially unwelcome, becomes fixed in the mind of the person over the course of time; a sudden, homicidal act, usually in the absence of any history of violence; and a postoffense period of relief during which memory is fully preserved for the event (4). There is usually no evidence of any conscious anger toward the victim, often an intimate or family member, yet the killing itself is a testament to the capacity for extreme aggression by the perpetrator.

There have been a number of case studies of catathymic homicide published during the past 60 years (5,6), although its rarity has precluded any large, comparative investigations. This is a case of catathymic infanticide, and to the author's knowledge, the first such study presented in the scientific literature.

[1]Clinical Professor of Psychiatry, University of California, San Diego, School of Medicine, San Diego, CA.
[2]Adjunct Professor, University of San Diego School of Law, San Diego, CA.
[3]Faculty, San Diego Psychoanalytic Institute, San Diego, CA.

Received 30 May 2009; and in revised form 13 July 2009; accepted 19 July 2009.

The Offense

Mr. L awoke on a warm spring Monday morning and realized that the day had come when he would kill his 13-month-old son. He had stayed in bed most of the morning, as he usually did, and his wife had left much earlier to drop one of his three sons in day care. The second son was four, and the oldest son, who was eight, had left for school.

Mr. L dutifully took his medication cocktail of olanzapine, bupropion, venlafaxine, and lorazepam—it seemed as if they were throwing medications at him now—and continuously thought about the desperate state he was in. He had not worked for a year, he slept most of the day, he could not function as a parent, his wife was increasingly angry at him, and the mental health professionals were not helping. He was hopeless, helpless, humiliated, and felt almost paralyzed by his disorders. He had been diagnosed with major depression since mid-adolescence and had been treated by the same psychiatrist for 20 years with mixed success. His major depressive disorder—he knew the diagnostic terms quite well—did not seem to respond to the pharmaceutical efforts of his doctor, and his belief that he might benefit from psychotherapy appeared to fall on deaf ears. His course of treatment had been medication visits about once a month, and four hospitalizations: two in his late adolescence and two within the past 6 months. His second diagnosis, obsessive-compulsive disorder, did not appear until a decade ago when he was unable to decide whether he had mixed the cement correctly for his bricklaying. His productivity in the masonry company that had employed him for 13 years collapsed, but they generously continued to keep him on the payroll for his health insurance coverage. The other symptom that emerged was his fear that he "would hurt someone." He would arrive at a stop sign and not be able to venture forth in his auto until he had checked the cross traffic multiple times, often sitting at the stop sign for 10 min. He was compelled to return to the stop sign hours later to make sure he had not caused an accident from which he

had fled. Although his wife continued to tolerate his symptoms and emotionally support him, she found some of the repetitive behaviors benign, such as checking to see if the boys' milk had spoiled; and some of the behaviors quite bizarre, such as visiting the neighbors' home to see whether a stairway had collapsed which he had built years earlier.

But the idea of killing was different. He had attempted suicide when he was 12 by swallowing his mother's psychiatric medication, and he had most recently gestured suicidally by taking a large amount of clonazepam, which precipitated his fourth hospitalization. When he was 20, however, he had been "born again" as a Christian, and actual suicide was no longer an option. He would be consigned to Hell if he did take his own life. When his first son was born, moreover, he began to worry that he would not choose Jesus as his Savior when he grew up, and when he died, he would also go to Hell. This concern surfaced with the birth of each of his three sons, and he contemplated killing them through suffocation—as he put it, both a fear and an urge to do so—during their first year of life to ensure that their innocence would take them to Heaven. His desire to suffocate was most apparent when he held the infant and had the thought of squeezing his face into his upper arm. He never told anyone of these urges, including his psychiatrist, and never acted on his idea. His wife, however, vividly remembers occasions when he would be intensely concerned if he saw another infant with a blanket too close to his face.

During his third hospitalization 6 months before the homicide, a nurse recorded for the first time his desire to kill his youngest son: "complains of homicidal thoughts toward 9 month old son for 3 min last night. Depression. Feels hopeless." His worries about hurting someone were becoming more detailed and focused, and 2 months before the homicide he told his treating psychologist, "I've had thoughts of killing my son and going to prison." This time the target was his middle child. The psychiatrist and spouse were immediately informed, and a contract was signed by Mr. L agreeing that he would tell his wife and his professional treaters if he had any thoughts of hurting someone. There was also agreement that he would not be left alone and responsible for the children.

A new psychiatrist was now treating Mr. L after his fourth hospitalization 3 months before the homicide, and the diagnosis was changed to schizoaffective disorder. Personality disorder was also diagnosed for the first time in 20 years. Medication dosages were increased to no avail, and Mr. L continued to be socially withdrawn from both his family and friends, and virtually bedridden. ECT was considered, but Mr. L refused this treatment alternative.

The frequency and intensity of the fixed idea to kill his youngest son hounded Mr. L through the month of May. Although he did not tell anyone of these continuous thoughts, and he was resisting them less, his father was concerned enough to accompany him to his last psychiatric visit 18 days before the killing. It was a medication management visit, and the psychiatrist wrote in his note, "looking better to me, starting to function better. Brighter affect, more animated, appropriate, no thought disorder or suicidal ideation. Here with his dad, they both think he's doing better. Followup in 6–8 weeks." The psychiatrist did not ask him about any homicidal thoughts. His last psychotherapy session was 2 weeks earlier. He told his psychologist then, "My relationship with God is not terrific," and he did not return for any subsequent psychotherapy visits.

Mr. L got dressed, drew a tub of water, and drove the short distance to the day care where his two sons were. He lied to the young woman in charge, telling her he needed to take his youngest son to a doctor's appointment. He then drove home, making sure his infant was secure in his car seat. He took him to the bathroom and submerged him face down in the tub until he stopped struggling and was dead. He then called 911, told them he had drowned his son, and began administering cardiopulmonary resuscitation. When the police arrived, he initially lied again, telling them his son had fallen. He then confessed. They also found a handwritten note to his wife: "I'm sorry I killed John. I couldn't do anything else. I'm not a man. I am a coward. John is lying on the bed. Here is the key to the mower. Don't let the boys have it." When asked by the police why he had killed his son, he variously stated that he was confused, he did not know, he was depressed, and that it freed him of all the pressures in his life. What was most notable to the officers on the way to jail was his composure; they wrote that he seemed quite calm. His wife also noted his relief when she visited him a week later. That first evening in jail he asked God to forgive him and he slept well through the night.

Clinical Interview and Testing

Mr. L presented as a medium height, mildly obese, 37-year-old Caucasian male, neat, clean, cooperative, and eager to please 5 months after the homicide. Clinical interviewing was completely consistent with a diagnosis of major depressive disorder, recurrent, without psychotic features; and obsessive-compulsive disorder. He still reported intermittent symptoms while in custody, including a preoccupation that he would bite the television cord and a fear that someone would break out of jail because the brick mortar was not appropriately applied. He was being medically treated by the same psychiatrist who had seen him prior to the homicide and was taking venlafaxine 300 mg, bupropion 300 mg, olanzapine 10 mg, and lorazepam 2 mg as a prn for anxiety. He showed no evidence of psychosis and stated that he was "doing alright." There was no significant medical history other than asthma as a boy and a period of hypoxia when he was born.

Mr. L's history was negative for any antisocial behavior or conduct disorder prior to the crime, but he was medically treated for attention deficit hyperactivity disorder as a child. He had been an introverted boy growing up, largely ignored by his father who may have consumed alcohol excessively, and cared for by a mother who was clinically depressed. He was teased at school for his shyness and social awkwardness, and adapted by pleasing people and staying in the background. He was an average student with an average IQ. He completed 1 year of college prior to his first psychiatric hospitalization, and then did not return.

The maternal side of the family was significant for severe depression. Clinical interview and review of records confirmed that his mother had been clinically depressed since Mr. L was at least 6 years old, had been hospitalized for the first time at age 44, had at least two subsequent hospitalizations, and two trials of ECT with some positive effects. Her most acute period of psychiatric decompensation occurred at the same time her son became severely depressed in adolescence, and culminated in the parents divorcing and the mother abandoning the family to live elsewhere. Archival records indicated that her father had committed suicide "by hanging induced by worry" and her mother walked in front of a train 1 month later. Mr. L's mother was 5 years old at the time. There was also credible historical evidence that her grandfather committed suicide and her grandmother died in a mental hospital. Mr. L's two siblings, however, appeared to have escaped this genetic sword of Damocles.

Psychological testing, moreover, confirmed the findings of the clinical interview and the review of records. The Millon Clinical Multiaxial Inventory III suggested a personality disorder diagnosis

of Dependent Personality Disorder (BR 85) with additional avoidant (BR 84), schizoid (BR 83), depressive (BR 82), and masochistic (BR 78) features and traits. There were no elevations on the modifying indices. The Minnesota Multiphasic Personality Inventory-2 suggested a chronically anxious and depressed individual (72 Codetype) who naively attempted to put his best foot forward (L = 78). He evidenced low self-esteem (LSE = 75) and was hungry for reassurances (Dy = 68). To satisfy such neediness, moreover, he rigidly controlled and constricted his emotions (R = 81), particularly his angry and aggressive thoughts and feelings (OH = 69, ANG = 36, Sc2 = 78, Pa3 = 75). Both of these self-report measures suggested the intrapsychic conflict most central to his crime: he could not bear to risk abandonment by those he depended upon by allowing into his awareness—and God forbid expressed through his behavior—his anger and aggression. Instead, he turned his fury on himself, only to be projected outward when he held a helpless and dependent object, his infant son, in his arms.

The Rorschach provided additional psychostructural information. Although Mr. L produced a normative amount of responses (R = 25), he was highly defended against his own affects (L = 5.25) and evidenced a complete affective shutdown (FC + CF + C = 0). His attachment capacity and anxiety were normal (T = 1, Y = 1). Most notably he produced only one human detail response, an indication of his lack of whole object representations, a likely inability to mentalize (7), and the absence of empathy toward others (M = 0). He was bereft of any aggression responses (Ag = 0) and showed no expectations of cooperative interactions with others (COP = 0). The only significant clinical index elevation was a maximum score of 5 on his Coping Deficit Index, suggesting global deficiencies in social and interpersonal skills. His reality testing was within the normal range (X–% = 20, XA% = 80), and there were no indications of psychosis. There was no suggestion of chronic impulsivity (AdjD = 0).

Mr. L's PCL-R (8) score, a measure of psychopathy, was 4, indicating no evidence of psychopathic traits. There were also no indications on mental status exam of a need for further neuropsychological testing.

His recounting of the facts of the crime was very consistent with police reports and investigative findings. He stated that his first thoughts of homicide occurred a decade earlier, about the time of the onset of his obsessive-compulsive disorder. He found that his conversion to Christianity, and his born again experience, were initially quite moving, but he then began to worry that each subsequent son would make bad decisions when he grew up, not follow Jesus, and be consigned to Hell when he died. His first homicidal thoughts were documented 8 years before the killing in a psychiatric hospital note, "worries whether he will kill anybody or not." Such reports continued to intermittently appear in various psychiatric records *for a decade* until 3 weeks before the killing. The target of his homicidal fantasies moved from son to son, finally settling on his third and youngest. He deliberately chose to not tell anyone of his final decision to kill and reported that the thoughts occupied him all day long during the final weeks. The last homicidal thoughts he reported to anyone were directed toward his second oldest son 2 months before the killing. The specificity of his plan did not crystallize until the morning of the murder. After the homicide, Mr. L reported, "I was relieved at that time. It felt like a burden was lifted off of me. It was there for a couple of hours. Then at the jail I asked God to forgive me for killing me. I realized what I did. Not a lot of emotion. A little guilt. I felt God has forgiven me. I know by faith and his Word he will forgive us our sins and cleanse us from all unrighteousness. I slept that evening." His conscious motivations reported to me were attention-seeking, confusion, not knowing, and depression. His most plausible conscious motivation was that by killing one of his children, he "would be free of all the pressures in life," and his state of mind, which included feelings of desperation, humiliation for failing as a father and breadwinner, guilt, and paralysis, would be relieved.

The unconscious motivation was transference-based, and closely tied to his own mother's chronic depression. Very early in life, and for a number of years, Mr. L felt he was a burden to his mother and believed he was causing her suffering—a typical belief when a child has a mother who is chronically ill. By killing his own infant son, he was trying to annihilate the helpless dependency in himself. This reflects both self-loathing and guilt and is captured in his parapraxis (slip of the tongue) to the evaluating forensic psychologist, "to forgive me for killing me." His ability to mentalize—to reflect upon his own mind and the mind of another—was grossly impaired, and instead he projectively identified with, and hated, the dependency he witnessed in his own sons when they were infants.

The mental health professionals also failed him. His best treatment occurred during his four brief acute hospitalizations beginning in adolescence and extending over the next two decades. His outpatient treatment was inadequate for at least three reasons: (i) there was the assumption that medications alone, and brief monthly psychiatric medication consults, would be sufficient to address an individual with chronic familial depression, a personality disorder, and increasingly obvious failures to function at home and at work; (ii) there was no evidence that his last psychiatrist had reviewed the extensive psychiatric records from his past; and (iii) the last psychiatrist relied solely on the patient and his father's self-report during his last visit and made no inquiry concerning homicidal ideation or urges.

Mr. L's fundamentalist religion also contributed to the homicide. It provided him a religious sanction for the killing: if he murdered his son while an innocent, he could guarantee his ascension to Heaven when he died. Religious approval for homicidal violence, no matter how perverted the personal theology becomes, brings with it a resolve and the blessing of an ultimate authority—one who can also confer forgiveness once the act has occurred. The more fundamentalist the belief system, the more aggression can be sanctified, because unquestioning belief negates critical thought and typically breeds intolerance for others' opinions—and sometimes an utter disregard for their lives.

Mr. L pled guilty to the murder of his infant son. At his sentencing hearing, following the testimony of a forensic mental health professional and his spouse who continued to support him, Mr. L tearfully apologized to the community. He was sentenced to life in prison and is eligible for release in 25 years.

Discussion

This is a prototypical case of chronic catathymic homicide. The court-appointed psychiatrist who evaluated Mr. L 6 weeks after the murder wrote in his report, "I have done forensic evaluations for nearly 30 years…I have never seen a case quite like this…His reasoning does not explain an act that was totally out of character for him." That is the point. The motivational dynamics for catathymic homicide are often transference-based and not consciously understood by the perpetrator, the act is inexplicable by those who investigate it, and the prior absence of any violence completely obliterates the usual phenomenological explanations for intentional killing.

In chronic catathymic homicide, the choices are reduced, often in the dismal tunnel of depression, to either homicide or homicide/suicide. Other more reasonable choices cannot be considered, despite the lethal risk posed toward self or others. The finality,

narrowness, and intense drive of the catathymic process were captured most poignantly in this case when Mr. L left a note for his wife at the crime scene, "I'm sorry I killed John. I couldn't do anything else."

The chronic form of catathymic homicide is recognized by a lengthy incubation period, a sudden act of killing, and a period of relief (4). The victim is usually an intimate (mother, spouse, and in this rare case, a biological child). Acute catathymic homicides, on the other hand, are usually perpetrated against complete strangers, there is no incubation period, and the unconscious, transference-based motivation is usually uncovered by the painstaking elimination of all other motives for the crime (5). Although depression figured prominently in this case, it is usually not associated by professionals with risk for homicide, and the literature relating the diagnosis and intentional killing of another is scant (9–11). This case, however, illustrates once again the importance of inquiry concerning thoughts of *both* suicide and homicide in the evaluation of clinically depressed patients, particularly if their depression is treatment resistant. Most men who commit familicide are clinically depressed (11), and depression often figures prominently in males who commit mass murder (12).

This case also illustrates the contribution that religious belief can make to homicidal acts. Such beliefs are particularly relevant when they are unquestioned and absolute and provide a supreme being's sanction for the killing, at least in the mind of the perpetrator. This is superego aggression—the polar opposite of psychopathic aggression which is devoid of any internalized value—yet both can be planned, purposeful, and emotionless in their expression. The study of religious belief and homicide is beginning to receive some scientific attention through the scrutiny of extremely aggressive forms of martyrdom, such as suicidal-homicidal acts by Islamist true believers (13).

Most infanticides and child homicides are perpetrated by a family member and are the result of psychosis, intoxication, abuse, or neglect. None of these factors play a role in catathymia, hence the befuddlement of professionals who examine such cases.

Although individual studies such as this contribute little to scientific understanding because of their anecdotal nature, the accumulation of such studies often advance understanding through the aggregation of data. Catathymia, despite its discovery nearly a century ago, continues to be rarely understood and appreciated by forensic mental health professionals. Yet it remains an important concept for understanding homicides that first appear to be without motivation.

References

1. Wertham F. The catathymic crisis: a clinical entity. Arch Neurol Psychiatry 1937;37:974–77.
2. Maier H. Ueber katathyme wahnbildung und paranoia. Ztschr f d ges Neurol u Psych 1923;82:193.
3. Revitch E, Schlesinger LB. Murder: evaluation, classification, and prediction. In: Kutash IL, Kutash SB, Schelsinger LB, editors. Violence: perspectives on murder and aggression. San Francisco, CA: Jossey-Bass, 1978;138–64.
4. Revitch E, Schlesinger L. Psychopathology of homicide. Springfield, IL: Charles Thomas, 1981.
5. Meloy JR. Violent attachments. Northvale, NJ: Aronson, 1992.
6. Schlesinger LB. The catathymic crisis (1912-present): a review and clinical study. Aggress Violent Behav 1996;1:307–16.
7. Allen J, Fonagy P, Bateman A. Mentalizing in clinical practice. Washington, DC: American Psychiatric Publishing, 2008.
8. Hare RD. Manual for the psychopathy checklist-revised, 2nd edn. Toronto: Multihealth Systems, 2003.
9. Goldney RD. Family murder followed by suicide. J Forensic Sci 1977;9:219–28.
10. Rosenbaum M, Bennett B. Homicide and depression. Am J Psychiatry 1986;143:367–70.
11. Marzuk OM, Tardiff K, Hirsch CS. The epidemiology of murder-suicide. JAMA 1992;267:3179–82.
12. Hempel A, Meloy JR, Richards T. Offender and offense characteristics of a nonrandom sample of mass murderers. J Am Acad Psychiatry Law 1999;27:213–25.
13. Meloy JR. Indirect personality assessment of the violent true believer. J Pers Assess 2004;82:138–46.

Additional information and reprint requests:
J. Reid Meloy, Ph.D.
PO Box 90699
San Diego, CA 92169
E-mail: reidmeloy@gmail.com

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01487.x
Available online at: interscience.wiley.com

Commentary on: Bohan TL. President's Editorial—Strengthening Forensic Science: A Way Station on the Journey to Justice. J Forensic Sci 2010;55(1):5.

Sir,

The January 2010 issue of the Journal of Forensic Sciences (JFS) contains an article, "President's Editorial—Strengthening Forensic Science: A Way Station on the Journey to Justice" by Thomas L. Bohan (1). The article was read with the expectation of an exhortation to the American Academy of Forensic Sciences (Academy) to take the lead in promoting progressive change for equal justice, truth, and fairness concerning use of science in the courts.

The January 2010 JFS issue contains a bit of irony, which reflects the Academy's failure to confront issues within its own membership affecting equal justice, truth, and fairness. The President's Editorial is the first article (1). The issue's last article is a book review by Betty Layne DesPortes, on "Arson Law and Prosecution" (2). When read together, the first sentence of the editorial and last sentence of Ms. DesPortes' review are:

"In any legitimate justice system truth must play a paramount role and integral role...(t)he very survival of the rule of law depends not only on a justice system that administers the law fairly, but a system that is just by being well-grounded in truth."—"Laudable goal unmet."

The practice of forensic science and expectations of what it can accomplish, both positively and negatively, in our judicial system, have changed considerably during the last decade. However, its fundamental constructs have not. Forensic science is used to convict the guilty and to protect and exonerate the innocent. It is the most persuasive of all evidence (3). Legal rulings significantly influence how scientific evidence is presented in court, thereby affecting how it is used. Judicial decisions requiring scientific validation and proper methodology should reduce potential ethical lapses and fraud. Accordingly, these decisions will foster equal justice and promote fairness.

The Academy is premised upon advancement of good science in its application to the legal system. It does not exist to insulate, protect, or perpetuate the practice of bad science. However, too often, governmental agencies and laboratories engage in self-preservation through the Academy's auspices. One means of ensuring that the practice of bad science is eradicated through a properly functioning ethics enforcement system.

During the Bohan Presidency, several significant events occurred in addition to the publication of the National Academy of Sciences Report, "Strengthening Forensic Sciences in the United States: A Path Forward" (NAS Report) (4). The NAS Report addressed serious deficiencies in the nation's forensic science system and necessity for major reforms. Through President Bohan's guidance, the Academy's preamble was changed to include "integrity" and "competency," and two individuals were expelled for violating the strict ethical standards of the Academy. All of these events correlate with the Academy's 2010 program theme of "Putting Our Forensic House in Order: Examining Validation and Expelling Incompetence." Unfortunately, only the NAS Report was discussed in the editorial.

A primary concern of the Academy, however, must be in enforcement of ethical standards necessary to promote good science and protect the relevant legal community from the practice of bad science. Changes to the preamble and the enforcement of the Academy's Code of Ethics reflect the Academy's strong commitment to ethics, fairness, and the practice of good science. However, the Academy should not create a public impression of apathy or political discretion by failing to competently address ethics violations in a timely and effective manner.

Only time will tell whether Bohan's Editorial becomes a presidential hallmark. Presidential hallmarks are achieved through progressive change. Past presidents who have either attempted, or implemented, positive change through their office are few in number. All too often Academy presidents are lauded for maintaining the status quo. Bohan's Editorial will become a presidential hallmark if the Academy follows its directive to diligently promote and effectively implement enforceable standards and accreditation in forensic science. One possible means of facilitating this objective is to extensively distribute President Bohan's editorial, including but not limited to, judges (5), attorneys (6), politicians, law enforcement personnel, academicians, and anyone with an interest in forensic science.

President Bohan has attempted to put his "forensic house in order" through substantively confronting the challenges within the Academy and the NAS Report. The obvious need for practical solutions to facilitate ethical and utilitarian reformation within forensic science cannot be overstated. Based upon these achievements, the ensuing question is whether successive Academy presidents will maintain and develop the accomplishments initiated by President Bohan.

References

1. Bohan TL. President's Editorial—Strengthening forensic science: a way station on the journey to justice. J Forensic Sci 2010;55(1):5.
2. DesPortes BL. Review of: Arson law and prosecution. J Forensic Sci 2010;55(1):283.
3. *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579, 595 (1993).
4. National Research Council, The National Academies. Strengthening forensic sciences in the United States: a path forward. Washington, DC: National Academies Press, 2009.
5. One avenue for reaching judges is through the National Judicial College in Reno, NV.
6. Professional legal associations that should be approached to distribute the editorial are the American Bar Association, Washington, DC; the American Trial Lawyers Association, Dothan, AL.; the National Association of Criminal Defense Lawyers, Washington, DC; and the National District Attorney Association, Alexandria, VA.

Gil Sapir, J.D., M.Sc.
P.O. Box 6950, Chicago, IL 60680

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01458.x
Available online at: interscience.wiley.com

BOOK REVIEW

Leslie E. Eisenberg,[1,2] *Ph.D.*

Review of: *Handbook of Forensic Anthropology and Archaeology*

REFERENCE: Blau S, Ubelaker DH, editors. Handbook of forensic anthropology and archaeology. Walnut Creek, CA: Left Coast Press, 2009, 530 pp.

This comprehensive volume, the second in the World Archaeological Congress' *Research Handbooks in Archaeology*, takes a historical and world view of the disciplines of forensic anthropology and (forensic) archaeology as applied to crime scene documentation and investigation, disaster response, and crimes against humanity. Following a Foreword by the Series Editors (Hollowell and Nicholas) and an introduction by the editors, the book is divided into five discrete but complementary sections: (i) History of the Disciplines, (ii) Forensic Archaeology, (iii) Forensic Anthropology, (iv) The Crime and Disaster Scene: Case Studies in Forensic Archaeology and Anthropology, and (v) The Professional Forensic Anthropologist.

History of the Disciplines (nine chapters) provides the historical context for the fields of forensic anthropology and archaeology and reviews their respective regional foundations; at the same time, it examines how our practice and knowledge continue to evolve. *Forensic Archaeology* (two chapters) focuses on the search, excavation, and recovery of human remains, reviewing tried-and-true archaeological methodology (probing, coring, and limited excavation) as well as other less invasive geophysical techniques to locate buried remains.

The section entitled *Forensic Anthropology* includes 15 chapters on differentiating human from nonhuman bone, dating skeletal remains, analyzing commingled remains, assessing ancestry and the concept of race, sex, age, and stature estimation, and histological age estimation. Also included are chapters on ante- and perimortem trauma, forensic taphonomy, burned human remains, craniofacial identification (approximation and superimposition), forensic odontology, and forensic molecular (DNA) analysis.

The next section, *The Crime and Disaster Scene: Case Studies in Forensic Archaeology and Anthropology* (nine chapters), focuses on domestic homicide investigations, disaster response, victim identification, ethnic tensions, and crimes against humanity in the United States, United Kingdom, Asia, Bali, Iraq, Guatemala, and the former Yugoslavia.

Completing the volume, *The Professional Forensic Anthropologist*, presents six contributions that highlight ethical considerations, quantitative methods, legal aspects of identification, the expert witness and the courts, and the delicate balance that those who practice forensic archaeology for large bureaucratic organizations must sometimes achieve to maintain the integrity of evidence. The concluding chapter, written by Ubelaker and Soren, offers a recapitulation of the volume, highlighting the themes of search and detection, skeletal analysis, and professional conduct.

Other themes that are woven throughout the book are professionalism and accreditation and the increasing recognition by others of the value of death investigation as practiced by forensic anthropologists and archaeologists who apply their unique skills in medicolegal contexts.

The field of forensic anthropology has now become a global endeavor, lending a voice to those who can no longer speak for themselves and judicial rigor to the complex fields of human identification and death investigation. This superb volume brings together contributions from practicing professionals in Australia, Canada, France, Great Britain, Guatemala, Indonesia, Italy, Spain, South America, and the United States who "… often work in extraordinary and challenging social circumstances (Hollowell and Nicholas, p. 17)."

My only wish is that the *Handbook* text was printed in larger type, but that would have made for a volume double in size and prohibitively expensive. For those who may not be familiar with some of the contributors who practice outside of their scope (and geography) of interest, background on each contributor is presented. Images presented in this volume are not used gratuitously, and only to supplement and enhance each chapter.

Graduate students in anthropology, practicing forensic anthropologists, governmental and nongovernmental organizations, as well as private entities that engage forensic anthropologists and archaeologists in search, recovery, and human identification work at home and abroad, should read and own a copy of this book. This volume will also resonate with those in the professional forensic and archaeological community, and legal arena, who wish to learn more about the significant contributions that anthropologists make globally. For those individuals considering specializing in forensic anthropology, or applying their archaeological skills in the search, documentation, and recovery of human remains and evidence derived from nonarchaeological contexts, this book inserts a large dose of reality into what it really means to be an anthropologist first and a forensic anthropologist second.

[1]Wisconsin Historical Society, Madison, WI.
[2]Dane County Coroner's Office, Madison, WI.

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01461.x
Available online at: interscience.wiley.com

BOOK REVIEW

Eleanor A.M. Graham,[1] *Ph.D.*

Review of: *Handbook of Forensic Science*

REFERENCE: Fraser J, Williams R, editors. Handbook of forensic science. Devon, United Kingdom: Willan Publishing, 2009, 662 pp.

When agreeing to undertake a review of this book, I thought that it would be a relatively rapid and painless process. Half of this assessment was correct. Initial attempts to skim-read chapters immediately became many late nights of engrossed reading, and even more hours of contemplative thought, all of it very enjoyable. Although it is difficult to describe this book under the traditional definition of "handbook," the text within certainly delivers on the promises provided in the synopsis, reviewing forensic science in the wider scientific, political, economic, social, and legal contexts. In fact, it delivers more than it says on the tin.

The book is structured by four parts, each dealing with a particular aspect of forensic science. A useful and well-considered introduction is provided by the editors for each main part, which comments on the contents in the wider context of the handbook. Briefly, part one contains three subsections devoted to scientific practices for identifying individuals, identifying materials, and reconstructing events; each section is divided again into several chapters, covering in detail the dominant disciplines used in contemporary criminal investigations. In part one, and throughout this book, each chapter follows a similar format. A succinct history of the particular subject is provided, before details of contemporary methods are provided. The detail provided for practical methods is not, in general, sufficient for the reader to be able to perform the experiments described for most chapters in part one; hence the reticence to embrace the title of this book. This is not, however, a criticism. For those requiring this level of detail, each chapter is extremely well referenced, directing the reader to relevant texts and articles for each aspect of the practice being described. More importantly, the author of each chapter goes on to openly discuss the limitations and controversial aspects of their specialist field and provides some insight into future research and developments that will ensure the continued advancement of forensic science practice over the next few years.

Part two shifts the focus from the laboratory into the office, providing detailed information on how scientific resources are employed by the police to complement, but not replace, more traditional, qualitative methods of investigation. Included in this part is a detailed historical account of the formalization of forensic science provision in the United Kingdom, including commentary on political and social drivers for this change. Because of its impact and success, in this chapter and indeed the remainder of the book, discussions are heavily weighted toward DNA profiling. Aside from the chapter dedicated to the National DNA Database, it is evident that each individual author has made efforts to be as inclusionary of other disciplines as possible when commenting on recent, contemporary and future developments of policy and legislation effecting forensic science. Again, this is not a criticism, just recognition of the success that DNA profiling has enjoyed. Details are provided through well-written text of how the decision to employ scientific analysis in criminal investigations is made, and how effective this employment has been.

Part three deals with the critical outcome of forensic analyses, the weight of evidence. Although touched upon in the preceding text, the application of statistics to forensic science is dealt with in detail in this part. Again, the history of this application is neatly provided before the reader is walked through the critical statistical and probabilistic theories employed in the evaluation of scientific evidence. The next chapter is concerned with the terminology of the expert witness, highlighting the vast variety of words and phrases that are regularly encountered in expert testimony, with good commentary on the validity and appropriateness of each example. The role and expectations placed upon the forensic scientist are also discussed with reference to UK law in this part.

In the final part of this book, contemporary themes and debates in forensic science are highlighted, specifically international corroboration, ethics, and forensic expertise. It is this particular part of the book that most deviates from the term "handbook," yet it is extremely relevant and the most thought provoking of the entire book. In this part, the social and ethical impact of forensic science along with the necessity for continued professional development of forensic practitioners is outlined in persuasive terms before the book is concluded by the editors in a summative come optimistic chapter looking toward the future of forensic science.

In summary, this is a very well-written and extremely well-edited book, with contributions from many world renowned authorities on their subject areas. The style and content are easily accessible to readers at all levels of experience, from the interested lay person to forensic practitioners, police personnel, and lawyers. Enough information is provided in each chapter to satisfy many minds, but significantly the referencing provided throughout this book is extensive, yet directed, allowing the reader to effortlessly enhance their experience of a particular topic through further reading. It should be noted that, while inclusionary of worldwide practice when appropriate, this book is very much focused on the UK perspective. If the reader has a particular and specific interest in the historical, contemporary, and future practices of other geographical or political regions, they may find themselves slightly disappointed. I would, however, be surprised, given the relatively low financial cost and high quality of content, that any reader interested in forensic science, anywhere in the world, would be entirely disappointed by their purchase.

[1]East Midlands Forensic Pathology Unit, University of Leicester, Leicester, UK.

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01462.x
Available online at: interscience.wiley.com

BOOK REVIEW

Owen Middleton,[1] *M.D.*

Review of: *Color Atlas of Forensic Medicine and Pathology*

REFERENCE: Catanese CA, editor. Color atlas of forensic medicine and pathology. Boca Raton, FL: CRC Press, 2009, 424 pp.

In this compilation of photographs, the authors have made an admirable attempt to create a reference forensic pathology atlas. The text is also available in a DVD format for those more digitally inclined. Unfortunately, it is unclear who the intended audience for this publication is. The topics presented are those encountered in any general reference forensic pathology textbook, as are the majority of the included photographic subjects. Therefore, while the images are certainly interesting, this book is not likely to supplement the forensic pathologist's library. It has the potential to be an informative resource for those in training or simply interested in forensic pathology, but some of the photographs are not described with enough detail to allow for easy interpretation by someone without training in pathology.

The chapters are arranged in a logical order; however, this logic is somewhat undermined by the actual photographic layout of many of the pages, where the topics within a chapter occasionally seem to jump around or seem randomly inserted. Some figure descriptions are not clearly associated with their intended photograph, leading the reader to an erroneous initial impression until reviewing the remainder of the photographs and descriptions on the same page. The photographs generally provide decent examples of what the authors are trying to convey. However, many of the pictures are not of a quality one would expect to see in a new publication in this digital age.

Although one may anticipate typographical errors in any publication, more than a few in an atlas such as this is somewhat unforgiveable, as there are only 35 pages of true text in the chapters (excluding figure descriptions which are also not free of error). Especially concerning is an error in Chapter 8, within a description for gunshot wounds characteristic for a distant range of fire, in which the authors state "some experts prefer to call this *intermediate* [emphasis added] range, especially when the possibility of an intermediate target cannot be excluded." Assumingly, this was an erroneous use of "intermediate" instead of "indeterminate," but to the novice reader, it is incorrect terminology that comingles two ranges of fire. Another concerning statement, regarding acute alcoholism-related deaths, presents in Chapter 3 and states "death occurring from acute alcoholism is usually classified as *natural* [emphasis present in the published text], although this may not always be the case." While admittedly there may be office or regionally specific interpretations for classifying deaths, the statement conveys a certain finality of classification for this scenario, which by convention most would argue should actually be classified as an accident. Also, rare statistical statements, which are apparently taken from published studies, cannot be independently evaluated by the reader as no references are provided. While this book is certainly not intended to be an exhaustive review of the literature, if such studies are going to be included in the text, a reference should be provided. Dishearteningly, one sentence contributes to the common belief (regardless of whether it was intended to do so or not) that forensic pathologists actually work for the prosecution, by stating that during the pediatric autopsy, the pathologist must "exclude all other reasonable possibilities that undoubtedly will be introduced by the defense in the adjudication of the case." On the contrary, the pathologist must perform the job with due diligence while striving for accuracy, which sometimes includes anticipating salient questions from many interested parties, and *not* to appease one side of a potential debate.

In the preface of the book, the authors profess a desire to provide "precision and accuracy" in select forensic topics through the information and descriptions in this atlas, a goal which they do not ultimately achieve given the limited scope of the book, combined with layout flaws, scattered typographical errors, and some misleading or unfortunately worded statements. The book might still retain some value in a training program with the caveat that knowledgeable teachers must correct the misleading errors, as most of the information and photographs would be informative to those not already trained in forensic pathology. Perhaps future additions will account for the items that currently detract from this publication.

[1]Hennepin County Medical Examiner's Office, Minneapolis, MN.

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01463.x
Available online at: interscience.wiley.com

BOOK REVIEW

Mary K. Sullivan,[1] *M.S.N.*

Review of: *Forensic Nursing: A Concise Manual*

REFERENCE: Garbacz Bader DM, Gabriel S. Forensic nursing: a concise manual. Boca Raton, FL: CRC Press, 2010, 441 pp.

The authors of this text are both associate professors at Bryan-LGH College of Health Sciences, School of Nursing in Lincoln, Nebraska. Each has extensive experience in providing forensic nursing education as well as remaining active in forensic nursing practice. The Preface states that this text is intended to be an introduction into the world of forensic nursing science as well as a tool for reference for any member of a multidisciplinary investigative team. Educators and students alike should find this succinct manual useful for an orientation to the forensic nursing specialty, as well as an efficient way to review content for an examination on any subject matter presented in this book.

Within the 35 chapters, the authors have covered a wealth of topics, providing both basic information and education resources. The majority of the book is structured in outline form, distinguishing this text from competing works. In several chapters, charts, tables, line drawings, and black and white photographs have been incorporated to illustrate key points. The authors have also included contributions from other prominent forensic science experts to address selected topics such as crime scene investigation, deaths in nursing homes, excited delirium syndrome, profiling, and forensic anthropology.

Forensic Nursing: A Concise Manual addresses major content that a forensic nursing student would want or need to learn. It begins with historical information on the development of the specialty, roles and responsibilities, pertinent definitions, and educational guidelines. The authors then transition into the topics of death investigation, crime scene preservation, and evidence management. Attention is given to establishing the time of death and the many variables to consider as well as the forensic nurse's contribution. The roles and responsibilities of (nurse) death investigators and how this position interfaces with the medical examiner/coroner/forensic pathologist are also covered. The next several chapters address the nursing process, ethical considerations, forensic nursing, and the law and guidelines for providing expert witness testimony.

The manual then moves on to specific topics that are usually included in forensic nursing education and training. A chapter is devoted to providing the essential information needed for assessment and documentation on each of the following: blunt force injuries, sharp force trauma, gunshot wounds, asphyxial death, sexual assault, child abuse, and domestic violence. Additionally, the authors have included a much-appreciated chapter that quickly sums up the dos and don'ts for the emergency department clinician with regard to assessment of forensic scenarios, appropriate documentation, and basic evidence collection procedures.

There are chapters devoted to human trafficking and mass disasters. Certainly, a student would need access to other reading material on these forensic scenarios to gain enough background information to fully comprehend the scope of both of these important topics. Fortunately, lengthy bibliographies are provided at the end of each chapter for more research.

Guest authors wrote in the traditional paragraph format, and thoughts flow logically and completely. Contributed chapters contain more than adequate material to be an excellent introduction to each specialty topic. Chapter 8 – Crime Scene Investigation by Matthias Okoye – is written concisely and provides an excellent overview of all necessary elements of crime scene management. This is followed by a chapter that focuses on the forensic nurse's role at a crime scene. This chapter overlaps somewhat; however, it does include more detail on the first responder role and addresses scene management for the living victim as well as mass disaster. Gary Plank covers the subject of profiling in Chapter 34. He does an excellent job of providing the history behind this specialty as well as addressing criminal investigative analysis (CIA). The material is very thorough and clarifies misconceptions about this area of expertise. The references are complete and any student interested in this subject should come away with clear understanding of what profiling actually entails and how and when to utilize the CIA tool. Chapter 35 is written by Erin Kimmerle and addresses the field of forensic anthropology, a topic ordinarily not presented in basic forensic nursing texts. This is the largest chapter in book with 29 pages and several good quality between photographs and reference tables.

Chapters 19 and 20, written by Vincent and Theresa Di Maio, cover deaths in nursing homes and excited delirium syndrome. These were written in the outline format but included more than adequate detail and reference citations within the text. Each of these chapters could be an excellent resource for clinical forensic nurses working within a busy hospital setting.

The content was logically organized and facilitated prompt comprehension of important information. However, the majority of the text was presented in outline format. Although the information is accurate, readers would probably feel lost without some previous knowledge of the subject matter. The outline format provides an ideal tool for organizing class presentations for educators or for students to review content for an examination. This reader found that many excellent statements within the outline were teasers, stopping short of providing the essential details to guide clinical

[1]Clinical Forensic Nurse, Department of Veterans Affairs, Phoenix VA Healthcare System, Phoenix, AZ.

practice. Although a bibliography was provided at the conclusion of each chapter, it was frustrating that there are no precise citations within the text of the outlines. This is a serious limitation for those seeking in-depth information on certain points.

This relatively inexpensive book would serve well as a basic text manual or study guide for students or educators who want to organize key thoughts about forensic topics.

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01479.x
Available online at: interscience.wiley.com

BOOK REVIEW

Zachary D. Torry,[1] *M.D. and Stephen B. Billick,*[2] *M.D.*

Review of: *Evaluating Eyewitness Identification*

REFERENCE: Cutler BL, Kovera MB. Evaluating eyewitness identification. Oxford New York, NY: University Press, Inc., 2010, 149 pp.

Eyewitness testimony plays a critical role in criminal trials, as the testimony provided by witnesses makes a significant impact on the jurors and can contribute to a guilty verdict. However, there are myriad factors that may distort the memory and/or later identification made by the eyewitness and lead to an erroneous conviction. In *Evaluating Eyewitness Identification*, Drs. Cutler and Kovera do an excellent job of conveying information about eyewitness identification and the role of the forensic mental health professional in providing expert testimony on eyewitness memory.

In the first part of the book, "Foundation", the authors address the legal context, key concepts in eyewitness identification, and empirical foundations and limits. The first chapter, "The Legal Context", presents a useful description of the role that eyewitnesses play in the criminal justice system. Moreover, it provides a framework for the rest of the book by describing the impact of eyewitnesses on erroneous convictions, the safeguards of erroneous convictions, expert testimony on eyewitness memory, and the legal standards for expert testimony. The standards and case laws described, *Frye v. United States* (1923), the Federal Rules of Evidence (1975), *Daubert v. Merrell Dow Pharmaceuticals* (1993), *General Electric v. Joiner* (1997), and *Kumbo Tire Company v. Carmichael* (1999) are imperative for every forensic mental health professional to know. These standards are important for an understanding of the admissibility and the receptivity to expert testimony. The admissibility concepts – general acceptance, juror common understanding, prejudicial impact versus probative value, and juror confusion—are discussed throughout the book. This is of great assistance to a forensic mental health professional preparing for an admissibility hearing.

Identification procedures, factors determining the accuracy of eyewitness identifications, and the general concepts necessary for understanding the research underlying expert testimony are described in the second chapter, "Key Concepts in Eyewitness Identification." The actual research on general impairment factors, suspect bias factors, postdictors, and admissibility concepts are reviewed in chapter 3. The authors were quite comprehensive in their overview of the research methods used in research on eyewitness identification. The use of the scientific method and peer review add invaluable weight to this area of research and will provide substantiation in admissibility hearings. Of considerable importance are the discussions on the benefits and limitations on each type of methodology. There are concerns about the external validity of laboratory research of crime simulations on eyewitness memory, and, as later discussed, this is surely to come out during the cross-examination. Therefore, the benefits hold considerable weight and are essential to know. Perhaps, there could have been more focus on the benefits, as an aggressive prosecuting attorney will surely call into question the few studies showing that crime simulations are not always generalizable.

The figure in chapter 3, Factors in Eyewitness Testimony, should be memorized, and the focus of evaluation for every expert witness on eyewitness memory. In addition, the explanation of each factor and summary of research is invaluable. The witness confidence malleability and witness consistency were not only very fascinating but will be quite useful for a cross-examination challenging the admissibility based on common sense.

The second section of the book focuses on the actual application of the information and research on the evaluation, data collection, interpretation, and report writing and testimony. Chapter 4, "Preparation for the Evaluation." discusses the importance of staying current with original research as well as the ethical principles and specialty guidelines. This chapter provides a practical context for the earliest contact between the attorney and the expert starting with the initial consultation and progressing through the guiding discovery phase. Chapter 5 continues with the data collection, and chapter 6 deals with interpretation. At this point, the expert formulates his or her opinion and, with the attorney, plans the expert testimony. The authors describe how the interpretation process is best preformed – the facts of the case are compared against the research literature and the best practices in identification tests. Finally, the attorney will determine whether to offer expert testimony in the case. Chapter 6 provides an intriguing discussion as to how this determination is made.

The final chapter, perhaps the most applicable for the actual testimony, is "Report Writing and Testimony." This chapter suggests a useful outline for the expert's report and then goes into "Preparing for Expert Testimony." In this section of the chapter, the authors discuss the various roles of the expert and the attorney. The expert first edifies the attorney about the expert testimony; the attorney enlightens the expert about the courtroom culture and etiquette; and finally, the expert educates the judge and jury during the admissibility hearing and the actual trial. Then, a strategy is laid out for the direct examination, cross-examination, and redirect examination.

This book is highly recommended for any forensic mental health professional as it contains very valuable information and can serve as a guide to those who will be evaluating eyewitness testimony. This book is part of the "Best Practices in Forensic Mental Health Assessment" series that describes "best practice" as "empirically supported, legally relevant, and consistent" with

[1]Forensic Psychiatry Fellow, Department of Psychiatry, University of Pennsylvania, Philadelphia, PA.
[2]President of The American Academy of Psychiatry and the Law, Clinical Professor of Psychiatry, New York Medical College; New York, NY.

the applicable and professional standards. The series editors recommend reading the first book of the series, *Foundations of Forensic Mental Health Assessment*, for the general principles that apply to all types of forensic evaluations. This is best followed by the specific topical book to provide the reader the general principles as well as those that are particular to the specific assessment question.

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01480.x
Available online at: interscience.wiley.com

BOOK REVIEW

Christian Crowder,[1] *Ph.D., D-ABFA*

Review of: *An Introduction to Microscopy*

Reference: Bell S, Morris K. An introduction to microscopy. Boca Raton, FL: CRC Press, 2010, 164 pp.

The microscope has been an essential tool for many scientific disciplines since its conception in the 16th century. Despite major advances in microscope technology, the practical use of the microscope still requires knowledge of its mechanics and an understanding of the interaction between light and optics with the material being evaluated. Although the microscope remains a widely used instrument in the forensic sciences, the authors note that "microscopy as an academic subject is nearly a lost art." This should be concerning considering that the forensic scientist, in particular, must demonstrate the accuracy, precision, and repeatability of all analyses. To achieve this, one must appreciate the capabilities and limitations of the instruments used for analysis. *An Introduction to Microscopy* provides a much needed, focused volume that is justifiably offered as a text for academic courses or a stand-alone reference for professionals, with the understanding that the reader has previously obtained fundamental knowledge in mathematics and physics.

This 164-page volume contains nine chapters, each concluded with several critical thinking questions and easy-to-perform laboratory experiments with the microscope. As suggested by the title, this book is concise in that it provides the fundamentals of microscopy that are essential for understanding more advanced techniques and analyses. The two authors have considerable field experience in forensic casework and have developed notable academic programs in forensic science. Despite the authors' backgrounds in forensic science, the volume is not specifically written as a forensic text, making the content applicable to any field that utilizes a microscope. The written text is accompanied by 110 illustrations, six tables, and four appendices that contain suggested readings and resources, a glossary of terms and abbreviations, and answers to the questions located at the end of each chapter.

Following the introduction explaining the context and structure of the volume, Chapters 1 and 2 discuss the basic principles of light and matter and how light and optics interact. Chapters 3 through 5 describe the mechanics of the microscope and sample handling. More specifically, these chapters briefly discuss types of microscopes, the main components of the microscope, how to achieve correct illumination, and preparing, mounting, and measuring samples. Chapter 6 is dedicated to digital image capturing and optimizing image quality for photomicrography. The last three chapters are more advanced, explaining principles of polarized light microscopy, crystallography, and chemical microscopy.

Overall, the authors succeeded in designing a book that is concise and general, allowing it to be broadly applied to various disciplines. As just one example, Chapter 8 describes the interaction between crystalline materials and polarized light, which is of importance for those working in fields such as forensics, soil science, earth sciences, and pharmaceuticals. I believe the volume is an excellent stand-alone reference for professionals; however, the volume will likely require a more comprehensive companion text for an undergraduate course in microscopy and a discipline-specific companion text for upper level undergraduate and graduate courses. For example, there is no discussion regarding the comparison microscope and stereomicroscope, which are essential tools in the forensic disciplines. The authors do provide a caveat that the volume is not intended to cover all types of microscopy. It may be for this reason that the authors do not provide an illustration of an assembled microscope. Chapter 6, Photomicrography, is a must read for anyone that uses digital imaging for research or forensic case reports.

In my opinion, colleges and universities would benefit in adding *An Introduction to Microscopy* to their microscopy courses or any laboratory section of a course that utilizes the microscope. The organization of the book allows for the reader to flow from one chapter to the next, constantly building upon the previous chapter's subject matter. This easy to read, well-illustrated volume with practical exercises throughout the text would be a worthwhile addition to the libraries of practitioners in the forensic sciences.

[1]Forensic Anthropologist, Department of Pathology, Office of Chief Medical Examiner, New York City, NY.

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01481.x
Available online at: interscience.wiley.com

BOOK REVIEW

Max M. Houck,[1] *B.S., M.A., Ph.D. (ABD)*

Review of: *An Introduction to Crime Scene Investigation*

Reference: Dutelle AW. An introduction to crime scene investigation. Sudbury, MA: Jones and Bartlett, 2011, 532 pp.

In the Foreword to *An Introduction to Crime Scene Investigation*, Aric Dutelle, a former law enforcement officer and crime scene technician currently teaching in a criminal justice department at University of Wisconsin-Platteville states that despite several "notable introductory texts on crime scene investigation" being available, he felt they were all oriented to "those who are involved in some capacity in the field." Thus, Dutelle wrote this text "with the assumption that the reader is new to the field of crime scene investigation and crime scene processing." To that goal, Dutelle has certainly hit his mark. *An Introduction to Crime Scene Processing* is a very basic book, in my opinion suitable for high school science students or community college students with no science background. It contains some questionable material and a few errors but is otherwise suitable for a very basic course.

The chapters include an introduction, a chapter on the CSI effect (a nice addition), one on ethics (again, a good idea), and then chapters on first responders, personnel and safety, and crime scene methods. The book then goes on to the laboratory and covers various types of evidence (fingerprints, trace evidence, biologicals, blood stain analysis, impressions, arson, drugs, and digital). The final chapters deal with death investigations and special scenes (such as underwater).

A few of the author's choices strike off notes. For example, using neurolinguistic programming (eye movement in interrogations as a method of deception detection) as an example of a questionable method—is it really a *forensic* method? Surely more relevant examples could have been chosen. It also seemed contradictory to have a chapter on the CSI effect (essentially downplaying the media's take on how forensic laboratories really work) but use the hackneyed and salacious title "Ripped from the Headlines" for current case examples. Some of the figures are visually murky (Figure 9.5—too dark to see), some murky as to what they intend to show (Figure 9.7—is that a flower?), and some are not useful ("Figure 4.4 A first responder vehicle" shows a US Capitol Police SUV—hardly illustrative).

A few topics are glossed over (not surprising given the basic information offered) but some are plain wrong. For example, Dutelle states that hairs from individuals of various ethnicities "…exhibit microscopic characteristics that distinguish one racial group from another,"—it is not that simple or that exact. Also, and perhaps worse, Dutelle maintains that "…the sex of an individual is difficult to determine from microscopic examination [of human hairs]" (I would submit it is impossible) and that "…longer, treated hairs are more frequently encountered in female individuals" (apparently, the author, although at a university, has not ventured out among the modern student populace). Moreover, Dutelle offers the product rule as a valid method in the following example: "…if all eye witnesses agree that a white male suspect over the age of 30 committed a crime, then utilizing the product rule would lend statistical support to such matters, and explain the frequency of such an event occurring." He then uses CIA statistics in Table 9.1 to calculate that one in seven persons could have committed the crime, ignoring the lack of legal support for this type of calculation, starting with *People v. Collins,** not to mention the lack of statistical support. Using the same approach to estimate a vehicle's relative rarity (page 192) is perhaps less suspect but certainly not a statistically standard approach.

Other books on the market (Fisher's *Techniques of Crime Scene Investigation*, Gardner's *Practical Crime Scene Processing and Investigation*, Turvey and Chisum's *Crime Scene Reconstruction*, and Siegel's *Forensic Science: The Basics*, to mention a few) cover these topics better and, yes, in more detail. At a time when forensic science – to include crime scene investigation – is being taken to task for a lack of science, it hardly seems that a more basic book is what the professional forensic community needs. For high schools or community college classes, however, *An Introduction to Crime Scene Investigation* might just fit the bill.

[1]Director, Forensic Science Initiative (Research Office) and Forensic Business Development (College of Business & Economics), West Virginia University, Morgantown, WV.

**People v. Collins*, 68 Cal. 2d 319, 438 P.2d 33, 66 Cal. Rptr. 497 (1968).

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01482.x
Available online at: interscience.wiley.com

BOOK REVIEW

John V. Goodpaster,[1] *Ph.D.*

Review of: *Quantification in LC and GC: A Practical Guide to Good Chromatographic Data*

Reference: Kuss H-J, Kromidas S, editors. Quantification in LC and GC: a practical guide to good chromatographic data. Weinheim, Germany: Wiley-VCH, 2009, 358 pp.

Determining when and why quantitative chromatographic techniques are used in forensic science depends heavily upon the nature of the samples involved. For example, most traditional trace evidence units are more concerned with the presence/absence of a material. The exact composition of the material is not an issue and a rigorous quantitative analysis is not needed, and it is neither time nor cost effective.

In other cases, quantification is critical to the results of a forensic science laboratory. Examples include determining blood alcohol concentration by headspace-GC, purity determination of controlled substances by GC/MS, or quantifying controlled substances and/or their metabolites in bodily fluids using GC/MS and/or LC/MS. In "Quantification in LC and GC: a practical guide to good chromatographic data," edited by Hans-Joachim Kuss and Stavros Kromidas, the practical aspects of generating and analyzing chromatographic data for the purposes of quantification are discussed. At the heart of this book is that there are several contributors to the precision of a chromatographic system. These are injection, separation, detection, and integration. The focus of the book is how each of these factors (with a particular focus on the detection and integration of peaks) can manifest itself in chromatographic data.

The book is organized into three parts: Part 1 (Chapters 1 – 9) is concerned with evaluating chromatographic data, integration of peaks, simulating data, deconvolution, and data interpretation. Part 2 (Chapters 10 – 13) focuses on characterizing and evaluating data from specific chromatographic techniques such as GC, LC-MS, IC, and GPC/GFC/SEC. Lastly, Part 3 (Chapters 14 – 17) discusses the requirements for chromatographic data analysis according to the European Pharmacopoeia, U.S. Pharmacopoeia, and the FDA. A bonus CD is also included that contains various spreadsheets and data files that are discussed in the text.

Perhaps the most useful part of the book from the point of view of a forensic chemist would be Part 1. In particular, the fact that various instrument manufacturers utilize different integration algorithms is shown to have significant effects on the calculation of peak areas, and therefore sample concentrations. Furthermore, it is stressed that integration errors cannot be eliminated through the use of an internal standard. Ultimately, it is the nature of the chromatographic data itself (the number of peaks and extent to which they are resolved) that determines how best to treat the data. The default algorithms of a data system, particularly when a quantitative determination is legally relevant, should not be accepted lightly.

The discussion of GC-MS in Part 2 is a good overview of the instrumentation and concepts involved, but it does not offer much that would be new to an experienced practitioner. The Chapter on LC-MS was more useful as it discusses internal standards, matrix effects, and ion suppression and how they can affect the ultimate reliability of the technique. Part 3 is less relevant to a forensic audience as it focuses on the standards used for pharmaceutical analysis. It may be interesting, however, to those who are taking steps to certify their laboratories through organizations such as ISO.

Overall, the book is well organized, with key points highlighted in the text and summarized at the end of each chapter. Given that the majority of the contributing authors are from Germany, the language style and usage may be occasionally awkward for an American audience. The authors are also most concerned with laboratories that conduct environmental, clinical, or pharmaceutical analysis. However, given the increasing scrutiny of forensic science, the standards of analytical chemistry should be met in the protocols that are used routinely as a part of criminal investigations. The most relevant portions of this work, therefore, are those that apply to the treatment of chromatographic data in general (Part 1). It is this material that could be used by forensic personnel to evaluate whether their quantitative analyses are as reliable as they could be.

[1]Assistant Professor, Department of Chemistry and Chemical Biology, Forensic and Investigative Sciences Program, Indiana University Purdue University – Indianapolis (IUPUI), Indianapolis, IN 46202.

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01486.x
Available online at: interscience.wiley.com

BOOK REVIEW

Mark M. Pollitt,[1] *M.S.*

Review of: *Investigating Digital Crime*

REFERENCE: Bryant R, Editor. Investigating digital crime. Chichester, UK: John Wiley & Sons, 2008, 259 pp, ISBN 978-0-470-51601-0.

Since the latter part of the twentieth century, the computer has become the icon of modern society. A book combining crime and computers has great appeal. As a result, there have been many books written in the last few years that focus on that combination. This is good, as there are so many facets to this legal, social, and forensic phenomenon, and the technology and its use change virtually daily. Because this field is very broad, any author or editor who writes such a book must make choices, as did the editor of this book. These choices are what determine the value of the book to a particular reader. This review was written for a forensic journal, and I will focus my comments from that perspective.

The book's first chapter, grandly titled "The Challenge of Digital Crime," tries to frame the entire topic in a mere 24 pages. Given the short amount of space, it does a credible job, but it would have been even more effective in expanded form. The book then jumps directly into a chapter on the legal aspects of digital crime that focuses almost exclusively on United Kingdom law, and as such will be of interest only for readers who operate there or have an interest in British law.

The chapters on investigating and countering digital crime are also very short (29 and 17 pages, respectively) and while adequate, have a number of shortcomings. First, they mixed investigation and forensic examination of digital evidence in ways that would lead the reader to believe that these things were undifferentiated. Second, from a forensic perspective, they gave very little emphasis on good forensic practice, including only the Association of Chief Police Officers (ACPO) Good Practice Guide. While there is nothing wrong with the ACPO guide, it is very basic and the authors did not even offer any additional sources of forensic guidance. Throughout both of these chapters, the text and references are peculiarly anglocentric and do not take note of any international research, writing, or forensic organizations.

The remaining chapters focus on specific technologies and crimes such as encryption, networks, credit cards, intellectual property theft, money laundering, and child sexual exploitation. Many have short sections on conducting investigations concerning these crimes and/or technologies, but in insufficient detail to provide a practitioner with the actual ability to conduct these investigations. The last chapter, while again very short, is the most interesting and deals with the criminology of cybercriminals.

This book attempts two ambitious goals: broad coverage of digital crime coupled with brevity. It succeeded in these goals but at the cost of depth. On balance, this book will appeal to those who have a limited knowledge of digital crime and a specific interest in its investigation within the United Kingdom. With the exception of the last chapter, seasoned digital investigators will find little new material or guidance from this book. Forensic practitioners will be left wanting a much more detailed and focused treatment of digital evidence, its acquisition, examination, and analysis.

[1]Visiting Faculty, National Center for Forensic Science, University of Central Florida, Orlando, FL.