David A. Stoney,[1] *Ph.D.*

Reporting of Highly Individual Genetic Typing Results: A Practical Approach

REFERENCE: Stoney, D. A., **"Reporting of Highly Individual Genetic Typing Results: A Practical Approach,"** *Journal of Forensic Sciences*, JFSCA, Vol. 37, No. 2, March 1992, pp. 373–386.

ABSTRACT: This paper considers the interpretation of serological typing data as a problem in forensic science, as opposed to a problem in population genetics or statistics. Controversies arising in this area are partly due to an overly narrow perspective that ignores basic forensic science principles. After an initial discussion of the special problem that deoxyribonucleic acid (DNA) blood typing poses to forensic science, the three difficulties common to all the proposed interpretive methods are discussed. These are: predicting genotype incidence from allele frequencies, predicting frequencies for the joint occurrence of genotypes in a number of different genetic marker systems, and determining the appropriate population to use to measure the frequencies. The inability to test assumptions that are inherent in our routine methods is noted. This is a procedural weakness that unnecessarily limits the admissibility of DNA typing evidence in court. A practical solution to this problem is offered that begins with minimal assumptions. Initially a statement is made based on (1) how many reference samples the laboratory has typed and (2) how many of these samples show genotypes corresponding to the case samples.

The second stage of the presentation begins with a statement that additional assumptions are necessary to fully interpret the evidence and that although these assumptions are scientifically very reasonable, they cannot be absolutely proven. The presentation can then proceed, if desired, to consideration of the specific assumptions and frequency estimates of any of the methods that have been proposed to date.

To follow this approach population data must be kept in a form that allows the simple first-stage statement to be made. This means that each individual's record would include typing results in each genetic marker system. Although this method of data storage differs from that used in most forensic science laboratories, it is exceptionally versatile, and allows great flexibility in data analysis.

KEYWORDS: pathology and biology, serology, statistics, deoxyribonucleic acid (DNA), evidence interpretation, blood typing, database, population

Procedures for interpreting highly individual serological typing results are currently highly controversial. This controversy has arisen after hasty adoption of methods from other disciplines, specifically from statistics and population genetics. This paper takes a step backward from the controversies that have developed and considers the basic forensic science behind the problem.

Received for publication 6 May 1991; revised manuscript received 8 July 1991; accepted for publication 25 July 1991.

[1]Assistant professor and director of forensic sciences, Department of Pharmacodynamics (M/C 865), University of Illinois, Chicago, IL.

There are three difficulties common to all proposed methods for interpretation of highly individual genetic typing results. These difficulties are caused by:

—predicting genotype incidence from population-derived allele frequencies (usually involving an assumption of Hardy-Weinberg equilibrium);
—predicting frequencies for the joint occurrence of genotypes in a number of different genetic marker systems (usually involving an assumption of independence among the systems); and
—determining the appropriate population to use to measure the frequencies.

These three problems have been with us all along in forensic serology, but they have received exceptional attention with the advent of highly specific deoxyribonucleic acid (DNA) typing methods. The first two issues have become critical because of the inability to justify our traditional assumptions. With conventional forensic serology we were able to *test* assumptions of Hardy-Weinberg equilibrium and to *test* assumptions of independence among bloodgroup systems. These tests could be conducted by routine statistical methods on databases of reasonable size because of the relatively few genotypic combinations in these systems. With highly polymorphic DNA typing, and increasingly specific results, we can no longer directly test these assumptions. The approach offered in this paper begins by recognizing this limitation and presenting results without incorporating these assumptions.

A two-stage presentation of results is suggested. At the first stage a relatively simple statement is made based on (1) how many reference samples the laboratory has typed and (2) how many of these samples show genotypes corresponding to the case samples. The second stage of the presentation begins with a statement that,

> There is no definitive method for calculation of frequencies for highly individual genetic typing results. Although a number of scientifically reasonable methods have been proposed, each of these involves additional assumptions. These assumptions, although straightforward and reasonable, cannot be proven.

The presentation can then proceed, if desired, to consideration of the specific assumptions and frequency estimates of any of the methods that have been proposed to date.

To follow this approach population data must be kept in a form that allows the simple first-stage statement to be made: we must be able to determine how many persons in our population database match the genetic typing results found in the case. This means that each individual's record would include typing results in each genetic marker system. This method of data storage differs from that used in most forensic science laboratories, but it is exceptionally versatile, allows a variety of additional statistical treatments, and avoids implicit assumptions. It can also be easily adapted to give flexibility in defining relevant subpopulations.

The third issue noted above, determining the appropriate population to use, has received attention largely due to added scrutiny from attorneys and population geneticists. There are basic, inherent difficulties in defining relevant populations for forensic science applications. Neither the approach offered here, nor any other, can avoid these difficulties. Perhaps we will be able to reach a consensus on the best options, but whether or not that comes to pass, we need to understand the limitations that are inherent in our work. In this paper the limitations will be discussed and the advantages of flexible construction of databases will be emphasized.

It is stressed that commentary on the various options for statistical calculations is being explicitly avoided. Instead we consider the issues common to all *highly individual* genetic marker systems, or combinations of systems. Depending on the particular DNA typing methodology, interpretive procedures enjoy different levels of controversy. For poly-

merase chain reaction (PCR) typing of coding regions of DNA, for example, there are established databases on which assumptions have been tested to the satisfaction of many. There is much more debate concerning the interpretation where typing procedures cannot differentiate discrete alleles and questions of binned alleles versus pattern matching arise. We do not differentiate among these fundamentally different aspects of the present DNA typing problem because we propose a more basic solution to the initial interpretation problem based on simple empirical observation. The value of this approach and the general issues discussed here will be important whenever the lack of clear consensus on the methodology of statistical calculations threatens to prevent explicitly valuable evidence from being introduced into court.

Overall Perspective: The Subjectivity of Absolute Identity

We casually use the term "fingerprinting" to describe the individualization process to lay persons. Inherent in the term fingerprinting is the concept of individualization—absolute specificity and absolute identification. We aspire to this and it is our ideal in one sense: it is the result that we would like our testing to have. Fingerprinting is a convenient term and helps the lay person understand. But it is ironic and deceptive that we use this term when describing forensic DNA profiling, because of the fundamentally different process of evidentiary evaluation. We have recognized this to some extent, and have tried to substitute the term "DNA profiling" for most applications.

In this section a comparison is made between how fingerprint experts reach conclusions of absolute identity and how we are attempting to reach these same conclusions through genetic marker typing. Through this comparison the special nature of our discipline emerges and the need for a fresh perspective on our interpretive methods becomes clear. The reader who responds, "But fingerprint identification is completely different—we only use the term to make an analogy," is correct to some degree, but naive as well. The analogy is with the underlying *purpose* of the examination and this commonality of purpose transcends the more superficial descriptive role.

In fingerprint comparisons examiners note the details in the patterns of ridges. Beginning with a reference point in one pattern, a corresponding point in a second pattern is sought. From this initial point the examiner then seeks neighboring details that correspond in their form, position, and orientation. These features have an extreme variability that is readily appreciated intuitively and which becomes objectively obvious upon detailed study. With the sequential documentation of more and more corresponding features between two patterns, scientist and lay person alike become subjectively certain that the patterns could not possibly be duplicated by chance.

What has happened here is analogous to a leap of faith. It is a jump, or extrapolation, based on the observation of highly variable traits among a few characteristics, and then considering the case of many characteristics. Duplication is inconceivable to the rational mind and we conclude that there is absolute identity. This leap, or extrapolation, occurs (in fingerprinting) without any statistical foundation, even for the initial process where the first few ridge details are compared. Although founded on scientific observations, the process of comparison and the conclusion of absolute identity is explicitly a subjective process. The conclusions are accepted and supported as subjective—very convincing, undoubtedly valid, but subjective.

In contrast, with genetic marker typing we view our increasing evidential value as a stepwise process. We detect a series of traits, each one of which is, to some degree, rare. Next we estimate the chance of each trait occurring randomly in a population. To proceed we need to combine these rare events and derive an estimate for the frequency of their joint occurrence. Historically we have done this using simple statistical methods and applying basic principles of genetics. With the advent of DNA typing methods, however,

the frequencies of occurrence that we are estimating have become much much rarer. If we extrapolate the laws of population genetics and assume that they apply to our rare events, we do get very low probabilities. Unfortunately, with increasingly rare alleles and genotypes we are finding that we can't substantiate the validity of these laws using the sciences from which the laws were derived. This is at the root of the controversy between population geneticists and forensic serologists.

To understand this conflict we need to recognize that with the goal of individualization through forensic serology we are doing something new, something that will require special attention and development. Are we really trying to *prove* uniqueness using statistics and population genetics? The contrast with fingerprint comparisons is important. We hold fingerprint specificity and individuality up as our ideal, yet this is achieved only through a subjective process. In fingerprint work we become subjectively convinced of identity; we do not prove it. And this works just fine. For fingerprints.

Whether or not we reject the more subjective leap to identity, we must realize that to reach absolute identity, or its probabilistic equivalent, through an objective process is not possible. Probabilities are objective when they can be tested and reproduced. Probability estimates of one in a hundred billion (people) are not objective. They have no objective meaning. They mean that we are looking at something very very rare, maybe something we've never seen before, but we are not in an objective, real world when we speak of this probability. Substitution of pseudo-objective calculations, as noted by Kingston [*1*], encumbers, rather than aids, the legal process.

More specifically, the disciplines of population genetics and statistics will not lead us through an objective process to absolute identification. They cannot do this because as we approach very rare frequencies of events, or very rare alleles and genotypes, the strength of these disciplines wanes and their objectivity, so well-developed for *populations* and *multiple events*, gives way to subjectivity. The focus changes from general laws acting on populations or multiple events to one on specific events and the question of individual identity. We have something of a "Heisenberg's uncertainty principle" [*2*] here. The smaller and smaller the frequency we observe in statistics, the larger and larger the population we need in order to estimate it. To test correlations among various genotypes we need larger populations still. And when we have an issue that necessarily involves a local population, usually of uncertain, heterogeneous composition, we cannot test or reliably predict frequencies of these rare events and the interrelationship among these rare events. Eventually, we are faced with a choice between pseudo-objective calculations, made by extrapolating well beyond the provable, or accepting the reality of subjective certainty.

The lesson from fingerprint identification is that even without theoretical models and statistics we can, and do, make absolute identifications. We can apply expert, informed, scientific judgment and make the subjective determination of identity.

With DNA typing we are much better off than with fingerprints. We have disciplines that give us a tremendous start. Laws of genetics provide a foundation from which we can make reasonable inferences and critical judgments. Furthermore, we have tests that are separate and we have every good reason to believe that the results are statistically independent. But because we are better off we fear the subjective, and necessary, leap to individualization. We must either accept this leap or abandon the attempt to individualize blood. For DNA typing this will mean that our *provable* probabilities will be much, much more common than either our good science or common sense would allow.

As a practical matter the information content attainable through DNA typing methods is more than sufficient to establish absolute identity and even with excessively conservative statistics the variation will be sufficient for this purpose. Further statistical methods are also likely to be developed that eventually will be deemed acceptable by consensus standards. Although our difficulties will be eventually alleviated, the inability to achieve

absolute identity through objective methods will remain. As with the fingerprint examiners, we will have adopted a convention that we will accept as proof of identity.

Generation of Genotype Frequencies from Allele Frequencies

Historical Practice

The typical procedure for estimating the genotype frequency in forensic serology is to use allele frequencies, derived from population surveys, to predict genotype frequencies. Genotypes are the joint occurrence of a specific pair of alleles.

Beginning with the population survey, individuals are typed in a particular genetic marker system. Taking the phosphoglucomutase (PGM) system in a Caucasian population as an example, the survey by Grunbaum et al. [*3*] had the results

Type 1-1:	626
Type 1-2:	393
Type 2-2:	48
Rare types:	2
Total of 1069 individuals	

Omitting the two rare types found, leaves 1067 individuals divided among the three common genotypes. Allele frequencies p_1 and p_2 are next predicted from this genotype data, using the maximum likelihood estimator formulae:

$$p_1 = (2x + y)/2N \text{ and } p_2 = 1 - p_1$$

where x represents the number of Type 1 homozygotes and y represents the number of Type 1-2 heterozygotes. This results in PGM allele frequencies for Allele 1 and Allele 2 of 0.771 and 0.229, respectively.

Next the hypothesis of independence of the two alleles within the population is tested. That is, can we predict the genotype frequencies from the allele frequencies? This is also a test of the hypothesis of Hardy-Weinberg equilibrium. If the alleles are independent, then the frequencies for the three common genotypes will be well approximated by

$$\begin{aligned} 1\text{-}1 &= (p_1)^2 = 0.594 \\ 1\text{-}2 &= 2p_1p_2 = 0.353 \\ 2\text{-}2 &= (p_2)^2 = 0.052 \end{aligned}$$

Performing a chi-square test on these data results in accepting (failing to reject) the hypothesis of independence (see Table 1). Since the allele frequencies can predict the

TABLE 1—*Observed incidence of PGM genotypes in a population of 1067 individuals and their predicted incidence based on a hypothesis of independence.*[a]

Genotype	Observed Incidence	Predicted Incidence	Chi Square Value
1-1	626	634.27	0.108
1-2	393	376.78	0.698
2-2	48	55.95	1.130
		combined chi-square	1.936[a]

[a]The probability of obtaining the observed combined chi-square value under the hypothesis is approximately 0.16.

observed genotype incidence, they are then used for predictions of the random occurrence of the genotypes.

The Problem Posed by DNA Typing Methods

Compared with the conventional antigen and enzyme methods, DNA typing methods have much more polymorphism: there are many more alleles found in the population. We can still take a survey of individuals and determine allele frequencies, but it is much more difficult to test the hypothesis of random independence: there are many, many genotypic combinations. Many of these might not even be seen in a reasonably sized population. If we exclude rare alleles, as done in our example above, we exclude most or all of the survey. Our *habits* have induced us to take a parallel approach, however, and most laboratories keep only allele frequency data. An assumption of independence is then made to predict the genotype frequency. This practice is controversial because the assumption is not testable by conventional methods. New methods to test the assumption of independence are now being proposed [*4*,*5*] as are sophisticated methods that avoid the assumption of independence altogether [*6*]. Nothing that is said here is in conflict with these efforts and no doubt an accepted method to test this assumption will be developed, but our purpose here is to take a step backward, look at the forensic science problem, and gain perspective.

First, note the circular nature of our procedure. We survey genotypes, use the genotype frequencies to derive allele frequencies, and then use the allele frequencies to predict genotype frequencies. After this we feel lost because we cannot test the assumptions that allow the latter prediction.

The forensic science problem occurs case by case and is the estimation of the frequency for a *specific genotype*. We are getting distracted by the incomplete link between allele frequencies and genotype frequencies. We already had a measure of the genotype frequency, at least to a limited extent, when we surveyed genotypes. What is the need to use or test for Hardy-Weinberg equilibrium? The forensic science problem does not need this, and whereas it may be of some additional utility, it does introduce additional assumptions—assumptions that (at present) cannot be satisfactorily tested. It would be far better to avoid this dependence. To do so we need genotype frequencies, or at least a measure of genotype incidence in the population of interest. Before exploring some options, we will examine the problem of independence among genotypes in different marker systems.

Independence Among Genotypes in Different Marker Systems

Historical Practice

When using an assortment of different marker systems, independence is typically assumed and the individual frequency estimates for each system are multiplied together. This second independence assumption can be justified either directly using statistical tests, or theoretically based on proofs that the genetic loci responsible for the markers are not linked. This proof can be made either by mapping the genes to different chromosomes or through observations on pedigrees.

The Problem Posed by DNA Typing Methods

Direct statistical justification of the assumption of independence is not feasible when the number of genotypes becomes very large. Theoretical justification based on unlinked genetic loci remains, however, and has been seen as acceptable when homogeneous populations are being considered.

The major objection that has arisen is not attributable to the increased polymorphism of the DNA typing methods, but rather to the increased scrutiny placed on forensic science practices. This scrutiny has come, in turn, from the radically increased evidential value that has resulted from DNA typing methods. The objection is that for heterogeneous populations, made up of two or more ethnic groups, there will be an observable non-independence among the genetic markers. The objection is simply illustrated. Again using data from Grunbaum et al. [3], we begin with the observations of genotype incidence for esterase D (EsD) and haptoglobin (Hp) among Asian and black populations. These data appear in Table 2. Paired data, giving results for both Hp and EsD on each individual, were not presented and are not typically available in forensic science databases. What we will do here is compute genotype frequencies based on the above sampling and assume that *within each racial group* the two marker systems are independent. We can then calculate the expected frequencies of each of the nine Hp/EsD genotype combinations for the two racial groups. This calculation gives the expected frequencies shown in Table 3, assuming a population of 1000 for each race.

Taking a mixed population of 500 individuals from each racial group, we would have observed frequencies in each of the nine cells that are the average of the two values above. The values may then be tested for independence between EsD type and Hp type. For this population observed, expected and chi-square values are given in Table 4.

For this mixed population the overall non-independence of the two marker systems is easily demonstrated. This is the basis for criticism of routinely assuming independence among genetic marker systems. Responses to this criticism have taken several forms. A reasonable argument can be made that the overall discrimination of the methods is unaffected by even extremely heterogeneous populations. The work of Evett and Gill [7] is relevant to this point, even though it specifically examines the independence of allele frequencies. In their experiments artificially mixed populations were generated from populations with grossly different allele frequencies. They found little overall difference in the probability of duplication of a DNA type. When several typing systems are used, the non-independence in one system often corrects for non-independence in

TABLE 2—*Genotype frequencies of haptoglobin and esterase D in two populations* [3].

Blacks				Asians			
Hp (n = 460)		EsD (n = 770)		Hp (n = 1104)		EsD (n = 1428)	
Type	Frequency	Type	Frequency	Type	Frequency	Type	Frequency
1-1	135	1-1	626	1-1	148	1-1	561
1-2	236	1-2	140	1-2	444	1-2	674
2-2	89	2-2	4	2-2	512	2-2	193

TABLE 3—*Expected frequencies of haptoglobin and esterase D genotypes in populations of 1000 individuals under an assumption of independence.*

Hp Type	Blacks EsD			Asians EsD		
	1-1	2-1	2-2	1-1	1-2	2-2
1-1	238.6	53.4	1.5	52.7	63.3	18.1
1-2	417.1	93.3	2.7	158.0	189.8	54.4
2-2	157.3	35.2	1.0	182.2	218.9	62.7

TABLE 4—*Population-related nonindependence of Hp and EsD types; observed, expected, and chi-square values are given for each cell.*[a]

Hp Type		EsD Type 1-1	2-1	2-2
1-1	Observed	145.7	58.4	9.8
	Experiment	128.9	69.9	15.0
	Chi square	2.17	1.91	1.81
1-2	Observed	287.6	141.6	28.6
	Experiment	275.9	149.6	32.1
	Chi square	0.49	0.44	0.40
2-2	Observed	169.8	127.0	31.8
	Experiment	198.2	107.4	23.1
	Chi square	4.07	3.57	3.34

[a]The population is a mixed one of 500 blacks and 500 Asians using frequencies from Grunbaum et al. [3] and assuming independence of Hp and EsD type *within* each racial group. The probability of observing the combined chi-square value under the hypothesis of independence is approximately 0.001.

[b]Sum of chi-square values = 18.2; the P value for 4 degrees of freedom = 0.001.

another, and in any case the magnitudes of errors are seen to be of little significance *in the context of the forensic science question.* That is, a discriminating power of 1 in 10 billion is for all practical purposes the same as a discriminating power of one in 14 billion. Other responses to the criticism of population based non-independence are more appropriately discussed after looking at our third major topic: population issues.

Population Issues

A critical initial question that precedes all frequency estimation is the selection of the appropriate population for the given problem. Designating appropriate populations for forensic science problems is tricky [8]. There are basic, inherent difficulties in defining relevant populations for forensic science applications. Although there is no way to avoid these difficulties, we do need to understand them fully so that limitations which are inherent to our work are not attributed to faulty method or logic.

In forensic science we address the issue of identity by comparing two hypotheses: the hypothesis that a specific accused individual (suspect) committed a crime and the hypothesis that some other person committed the crime. The latter hypothesis requires assessment of the probability of the evidence occurring when a person is selected randomly from (whatever the) relevant population.

The relevant population is difficult to define. We wish to predict the frequencies of traits as they occur in randomly selected individuals—a field of individuals who are possible suspects. There will be some geographical region of relevance defined loosely by those who could have traveled to the crime scene to commit the crime. Insofar as there are geographical variations in genotype frequencies, sampling from a local population is desired. Furthermore, if the suspect was selected in part on the basis of a physical description, the population should be restricted to those fitting this description. Such a description might or might not be definite enough for superficial ethnic or racial classification.

Population databases of various types are available to forensic science laboratories, but none of these can be considered (strictly) random. We don't even know how to define the selection process or the geographical boundaries of the population. Practical

considerations limit forensic bloodtyping populations to those listed in Table 5. Case circumstances (or the question asked by the court) dictate which of these populations is most appropriate to use. Usually one wishes to best represent a person selected at random from a local population. Occasionally one wishes to restrict the question to a specific racial or ethnic group. This is done *not* simply because the accused is of this group, but *only* when there is independent evidence that the person who committed the crime is of a particular racial or ethnic group. This can occur in several ways:

—A witness or victim may give a description of the offender.
—The geographical circumstances may essentially restrict possible offenders to one group (for example, an offense on an island or an Indian reservation).
—Other physical evidence may indicate racial or ethnic group (for example, hairs or a particularly diagnostic genetic marker).

The ancillary evidence that restricts the racial or ethnic group of the offender needs to be examined critically. In the case of a description by a witness, the evidence is subjective and based on physical appearance. Ideally we would use a database of individuals who were selected using this same, subjective criteria. A database from an anthropological study of strictly defined racial groups would be less appropriate. It should be clear that although approximations and good judgments in data selection can be made on a case-by-case basis, there is no generalized best method. The reality of available data *and our inability to precisely define the desired data* make a strictly relevant population database unattainable.

In practice we can address this problem in several ways. The simplest is to select genetic marker systems where genotype frequencies are insensitive to the population variations. We can eliminate concern over geographical and racial variations if the traits that we measure show the same frequencies in these different populations. This is a good approach, but once again the highly polymorphic DNA typing systems limit our ability to fully test the hypothesis. Individual genotype frequencies are so low that variations between populations cannot be easily tested, given realistic population sizes.

A second approach to the problem, proposed by Buckelton and Walsh [9], is that of population stratification. This approach involves proving or assuming independence within each of various homogeneous population groups. This can be justified using the basic principles of genetics and demonstrating through family studies that loci controlling the various genetic markers are not genetically linked. One's local population is then considered, and using census data the proportion of each of the homogeneous population groups within the overall population is determined. To get the overall estimate for frequencies in the heterogeneous population, frequencies in each homogeneous group are weighted by their proportion of the total population. Obviously the success of this

TABLE 5—*Types of population databases available to forensic science laboratories.*

FROM IN-HOUSE TESTING
Casework samples (liquid blood from victims, suspects and paternity cases)
Staff samples
Post mortem samples (autopsies)
Samples from clinical laboratories
Samples from research projects (for example, twin studies)
FROM OTHER SOURCES
Data from other forensic science laboratories
Literature data

method is dependent on identifying subsets of one's population that correspond to homogeneous populations. Some populations will be mixed together and will defy even approximate ethnic classification. In our frustration with this issue we must remember that the difficulty lies with the problem as posed, rather than with our logic or methodology.

Requirements for a Comprehensive Database

To allow flexible choice of interpretive methods and to ensure no data loss, a database must meet the criteria listed below. Key elements of a database that allows this flexibility are summarized in Table 6.

1. Genotype frequencies must be available directly, rather than only as derived from allele frequencies. This is necessary to eliminate the routine need for the assumptions inherent in this derivation.
2. Typing results for each individual must be retrievable for the full set of blood group systems tested. This is necessary to eliminate the routine need for assumptions of independence among the blood group systems and for the testing of any correlations in complex populations.
3. The method used and test conditions must be retrievable for each system. This is necessary to accommodate changes in methodology and to allow tests of whether these changes affect the distribution of the data.
4. The source of each individual sample must be retrievable. This is necessary to keep track of geographical origin and for flexibility in selecting subsets of the data that may be appropriate for particular cases.
5. Any data available on the ethnic or racial classification of the sample, or on the appearance of the person, should be retrievable. This is necessary in order to keep track of subpopulations and to test for the effects of population stratification. Alternatively it needs to be known if the sample was meant to be "random," or was selected without reference to appearance or race.
6. A unique identifier of each individual is necessary in order to avoid duplicate entries into the database.

TABLE 6—*Key elements of records for a comprehensive database of typing results in forensic serology.*

Sample Description Fields
1A—Unique identifier for the individual
1B—Racial or ethnic indicator
1C—Physical appearance indicator
1D—Source of sample (type of survey, geographical origin)
1E—Case number or other cross-referencing information
Serological Typing Result Fields
System 1
2A—Typing result for Sample 1A in blood group System 1
2B—Method used for typing in System 1 (date, procedure keys)
2C—Special information relevant to System 1 (for example, gel identifier and lane number on a gel-based system)
System 2
2D—Typing result for Sample 1A in blood group System 2
2E—Method used for typing in System 2 (date, procedure keys)
2F—Special information relevant to System 2 (for example, gel identifier and lane number on a gel-based system)
System 3
. . . and so forth for each blood group system

A Practical Approach for the Reporting of Highly Individual Genetic Typing Results

The database described above is designed to allow full flexibility in the choice of interpretive methods. Specifically, it avoids linking the data storage to any particular assumptions that might be used in the interpretation stage. What follows is offered as a starting place for interpretation of blood typing evidence.

The most simple statement of results after a series of blood typing tests is that patterned after Marris's interpretation of glass evidence [*10*]. Marris found a match in several physical properties between glass from a broken crime scene window and glass fragments found in connection with a suspect. He then surveyed 64 glass objects and found that none of these other 64 samples agreed with the glass found in his case. He relates the following:

> The Judge asked, during the trial, what would be the chances of two pieces of glass, selected at random, agreeing. . . . Although it might be possible to calculate the theoretical probabilities, it was considered that as the effects of ingredients, manufacture and distribution of glass could not all be traced, an estimate of the chances would not be justified. The definite evidence was, that of 65 samples from various sources examined, only one (that from the [suspect]) agreed . . . with that from the window.

The analogous "definite evidence" for our serological testing is simply the statement that, "I tested the samples in this case in a variety of blood group systems and found them to agree with one another. In our database of N samples from various sources, no other samples show these same blood group types." Of the issues raised regarding forensic serology interpretation, this simple descriptive statement raises only two:

—How reliably were the samples compared? (That is, what are the match criteria and are they being consistently applied?)
—Are the samples in the database representative of an appropriate population?

There are no means to avoid these two fundamental issues, but all further assumptions are avoided. There is no assumption of independence, Hardy-Weinberg equilibrium, or anything else. Although obviously not exploiting the full value of the evidence, this is a good place to start because it is basic and defensible. A second state of interpretation can then be presented, beginning with a statement that

> There is no definitive method for calculation of frequencies for highly individual genetic typing results. Although a number of scientifically reasonable methods have been proposed, each of these involves additional assumptions. These assumptions, although straightforward and reasonable, cannot be proven.

The presentation can then proceed, if desired, to consideration of the specific assumptions and frequency estimates of any of the methods that have been proposed to date.

A natural next step in the interpretation is to assume a binomial distribution for the combined occurrence of the blood types seen in the case. Using this assumption an upper bound for the frequency of the combination of genotypes can be computed. Again, the assumptions of independence and Hardy-Weinberg equilibrium are avoided.

The procedure uses a numerical solution for confidence intervals when the sample size is large. The upper confidence limit for the proportion of samples having the corresponding type is computed from Eq 1 [*11*]

$$\mathrm{UCL} = \frac{(r + z_\alpha^2/2) + \sqrt{(r + z_\alpha^2/2)^2 - (N + z_\alpha^2)/N}}{N + z_\alpha^2} \qquad (1)$$

where N is the number of samples in database that have typing results in the systems that correspond; r is the number of corresponding samples observed in population database; and z_α is the normal deviate exceeded with probability of α. Table 7 gives upper bounds for the population frequencies for 95%, 99%, and 99.9% confidence with sample sizes of 100, 500, 1000, 5000, and 10 000 and observed frequencies in the sample of 0, 1, and 2 (see the Appendix for a sample calculation).

Another fairly simple extension of the basic statement has been proposed by Sudbury [*12*], following procedures similar to estimates of discriminating power and Gaudette's calculations of average probabilities. In his method a general statement would be made regarding the probability of a match of two randomly selected samples. Given a database of N samples, these would be intercompared, giving a total of $N(N-1)/2$ sample pairs. If m matching pairs were found, there would be an average probability of $2m/N(N-1)$ for a match of two randomly selected samples. Upper bounds for this probability could be set in a manner analogous to that presented above. Limitations of average probabilities have been discussed by a number of authors [*13–15*].

A simple initial approach, incorporating a minimum of assumptions, seems an appropriate place to start. The particular merits of these methods are not so much the issue here as are three specific points:

1. To calculate even these basic statistics, a database is needed that preserves genotypic data and that links the results of one individual in the various typing systems.
2. In the *initial* assessment of the evidence avoiding complex interpretive assumptions will make the presentation of the evidence more robust.
3. If the court allows additional assumptions, or if a particular jurisdiction settles upon a standard interpretive methodology, the database used by the laboratory should allow rapid and flexible data retrieval.

It is *expected* that from this beginning we would normally adopt further assumptions and proceed with additional calculations. Unfortunately, the choice of these further operations is currently highly controversial and has been the basis for exclusion of DNA blood typing evidence in court. It is our hope to separate the basic blood typing data and methodology from a controversy that lies in the choice among alternative interpretive methodologies. The data, to some degree, speak for themselves. The message can be

TABLE 7—*Upper bounds for population frequencies computed from Eq 1 for various sample sizes and observed frequencies.*

	Sample Size, n				
r	100	500	1000	5000	10 000
			95% CONFIDENCE		
0	0.026 35	0.005 383	0.002 699	0.000 540 9	0.000 270 5
1	0.043 59	0.008 915	0.004 470	0.000 896 1	0.000 448 2
2	0.058 65	0.012 02	0.006 026	0.001 208	0.000 604 2
			99% CONFIDENCE		
0	0.051 32	0.010 70	0.005 381	0.000 108 1	0.000 540 7
1	0.068 92	0.014 39	0.007 233	0.001 453	0.000 726 8
2	0.084 80	0.017 73	0.008 913	0.001 791	0.000 895 9
			99.9% CONFIDENCE		
0	0.087 16	0.018 74	0.009 458	0.001 906	0.000 953 9
1	0.104 5	0.022 49	0.011 35	0.002 288	0.001 145
2	0.120 6	0.025 98	0.013 12	0.002 644	0.001 323

clarified substantially and validly through statistical models, but additional assumptions are necessary. When other scientists or legal practitioners are prepared to accept these assumptions, we can provide a reasonable (and better) interpretation of evidential value. If the assumptions are disputed, and we are prevented from making them, we should still be able to present an assessment of evidential value. Although the methods suggested here will grossly understate the true evidential value, they have minimal risk of outright exclusion of the evidence.

Acknowledgments

I wish to thank the South Australian Forensic Science Centre and staff for providing the opportunity and environment for the development of this work. I also wish to thank Thomas A. Kubic and JoAnn Buscaglia for their help.

APPENDIX 1. Example of an Upper Bound Calculation using Equation 1

Suppose that, in a sample size of 500, we observe a characteristic once. At 99% confidence the three equation parameters are:

$$r = 1$$
$$N = 500$$
$$z_{99} = 2.326 \qquad z_{99}^2 = 5.410$$

and Eq 1 evaluates to:

$$\frac{1 + 5.410/2 + \sqrt{(1 + 5.410/2)^2 - (500 + 5.410)(12)/500}}{500 + 5.410} = 0.014\,39$$

References

[1] Kingston, C. R., "The Law of Probabilities and the Credibility of Witnesses and Evidence," *Journal of Forensic Sciences*, Vol. 15, No. 1, Jan. 1970, pp. 18–27.

[2] "Heisenberg's Uncertainty Principle," *Encyclopedia of Physical Science and Technology*, R. A. Meyers, Ed., Vol. 10, Academic Press, New York, 1987, p. 518.

[3] Grunbaum, B. W., Selvin, S., Myhre, B. A., and Pace, N., "Distribution of Gene Frequencies and Discrimination Probabilities for 22 Human Blood Genetic Systems in Four Racial Groups," *Journal of Forensic Sciences*, Vol. 25, No. 2, April 1980, pp. 428–444.

[4] Odelberg, S. J., Plaetke, R., Eldridge, J. R., Ballard, L., O'Connell, P., Nakamura, Y., Leppert, M., Lalouel, J.-M., and White, R., "Characterization of Eight VNTR Loci by Agarose Gel Electrophoresis," *Genomics*, Vol. 5, 1989, pp. 915–924.

[5] Delvin, B., Risch, N., and Roeder, K., "No Excess of Homozygosity at Loci Used for DNA Fingerprinting," *Science*, Vol. 249, 21 Sept. 1990, pp. 1416–1420.

[6] Berry, D. A., Evett, I. W., and Pinchen, R., *Journal of Royal Statistical Society (C)*, 1991, in press.

[7] Evett, I. W. and Gill, P., "A Discussion of the Robustness of Methods for Assessing the Evidential Value of DNA Single Locus Profiles in Crime Investigation," *Electrophoresis*, Vol. 12, No. 2–3, Feb./March 1991, pp. 226–230.

[8] Stoney, D. A., "Transfer Evidence," *The Use of Statistics in Forensic Science*, C. G. G. Aitken and D. A. Stoney, Eds., Prentice Hall, London, 1991, in press.

[9] Buckelton, J., Walsh, K., Seber, G., and Woodfield, D., "A Stratified Approach to the Compilation of Blood Group Frequency Surveys," *Journal of the Forensic Science Society*, Vol. 27, No. 2, March/April 1987, pp. 103–112.

[*10*] Marris, N., "Identification of Glass Splinters," *The Analyst*, Vol. 59, 1934, pp. 686–687.
[*11*] Crow, E. L., Davis, F. A., and Maxfield, M. W., *Statistics Manual*, Dover, New York, 1960, pp. 51–52.
[*12*] Sudbury, A., discussions at the Canberra Meeting on the Feasibility of a National [Australian] DNA Profiling Database, Jan. 15–18, 1991.
[*13*] Aitken, C. G. G., "Average Probabilities—In the Absence of Frequency Data is There Anything Better?" *Journal of the Canadian Society of Forensic Science*, Vol. 20, No. 3, Aug. 1987, p. 287.
[*14*] Gaudette, B. A., "Evaluation of Associative Physical Evidence," *Journal of the Forensic Science Society*, Vol. 26, No. 3, May 1986, pp. 163–167.
[*15*] Stoney, D. A., "Evaluation of Associative Evidence: Choosing the Relevant Question," *Journal of the Forensic Science Society*, Vol. 24, No. 5, 1984, pp. 473–482.

Address requests for reprints or additional information to
David A. Stoney, Ph.D.
Department of Pharmacodynamics (M/C 865)
University of Illinois
Box 6998
Chicago, IL 60680