J Forensic Sci, Mar. 2005, Vol. 50, No. 2
Paper ID JFS2004206
Available online at: www.astm.org

Anne H. McNamee,[1] *D.D.S., M.Sc.; David Sweet,*[2] *D.M.D., Ph.D.; and Iain Pretty,*[3] *B.D.S., M.Sc., Ph.D.*

A Comparative Reliability Analysis of Computer-Generated Bitemark Overlays*

ABSTRACT: This study compared the reliability of two methods used to produce computer-generated bitemark overlays with Adobe Photoshop® (Adobe Systems Inc., San Jose, CA). Scanned images of twelve dental casts were sent to 30 examiners with different experience levels. Examiners were instructed to produce an overlay for each cast image based on the instructions provided for the two techniques. Measurements of the area and the x–y coordinate position of the biting edges of the anterior teeth were obtained using Scion Image® software program (Scion Corporation, Frederick, MD) for each overlay. The inter- and intra-reliability assessment of the measurements was performed using an analysis of variance and calculation of reliability coefficients. The assessment of the area measurements showed significant variances seen in the examiner variable for both techniques resulting in low reliability coefficients. Conversely, the results for the positional measurements showed no significant differences in the variances between examiners with exceptionally high reliability coefficients. It was concluded that both techniques were reliable methods to produce bitemark overlays in assessing tooth position.

KEYWORDS: forensic science, forensic dentistry, bitemarks, overlays, reliability

Courts have recently taken an aggressive approach toward the scientific foundation of expert testimony (1). The validity and reliability of scientific techniques used in the courtroom have brought many previously accepted methods of forensic investigation under closer inspection. Significant cases such as *Daubert* (*Daubert* v. *Merrell Dow Pharmaceuticals*, 509 U.S. 579, 1993) and *Kuhmo* (*Kuhmo Tire Co*. v. *Carmichael*, 526 U.S. 137, 1999) have demonstrated that scientific evidence must meet a minimum level of judicial scrutiny before testimony is accepted.

Although not completely objective, the use of computers is forging the way toward less subjectivity in the comparisons of bitemarks. With advancements in computer technology and imaging, a movement is underway to integrate the numerous digital applications that are beneficial in a forensic investigation to bitemark comparisons.

Bitemark investigation includes the physical and metric analysis of a suspect to a bite injury. The physical comparison of the suspect to the bite is achieved by producing a transparent bitemark overlay. This aids in visualizing the biting edges of the teeth while they are compared to the bitemark photograph. An overlay contains outlines of the perimeters of the suspect's anterior teeth that would most likely be seen in a resulting bitemark. This outline produces what is referred to as a hollow-volume overlay. Because all of the edges of the anterior teeth may not be in the same plane of occlusion, a complete physical analysis would additionally include the assessment of the suspect's dental casts. These differences should be considered when producing an applicable overlay.

Numerous techniques have been developed to produce bitemark overlays. Several studies have evaluated the accuracy of various overlay production methods. In 1998 Sweet and Bowers compared the five most common overlay methods (2). Data were collected for the area and relative rotation from each tooth present on the 150 sets of overlays studied. Results of a multivariate analysis of variance showed significant differences between the techniques for both area and rotation. The computer-based production method was the most accurate and most objective, and was used as a gold standard to compare the accuracy of the other methods.

Adobe Photoshop® is a popular software program used to select the biting surfaces of anterior teeth and produce an overlay (3,4). In 2001, Pretty and Sweet examined the effectiveness of the use of bitemark overlays produced using Photoshop® (5). This study established error rates for the use of overlays and laid a foundation for further investigation.

Overlays are also produced using software programs that analyze the tonal contrast found in scanned cast images. Using tone-line imaging, it has been suggested that artistic ability, knowledge of dental anatomy, and personal bias do not influence the result producing a more objective analysis (6). Naru and Dykes developed a technique that enables an overlay to be created by inverting the tonal elements within the cast image to produce an outline of the original biting edges (7,8).

These two methods are the most popular for the production of computer-generated overlays. They are used widely in North America and Europe, respectively. In the current study, we will compare the results of each technique to evaluate their comparative reliability.

Methods

Participants were asked to produce hollow-volume overlays using two different computer techniques. The following groups were selected to address this issue: a) Diplomates of the American Board

[1] Bureau of Legal Dentistry, University of British Columbia, Vancouver, BC, Canada.
[2] University of British Columbia, Director, Bureau of Legal Dentistry, Vancouver, BC, Canada.
[3] University of Manchester, Unit of Prosthodontics, Department of Restorative Dentistry, Manchester, England.

* A portion of this work was presented at the American Academy of Forensic Sciences Annual Meeting in Dallas, TX on February 20th, 2004. Financial support was gratefully received from the American Board of Forensic Odontology, the American Society of Forensic Sciences, and the University of British Columbia Faculty of Dentistry.
Received 18 May 2004; and in revised form 18 Aug. 2004; accepted 3 Oct. 2004; published 2 Feb. 2005.

of Forensic Odontology with extensive bitemark experience, b) Forensic Dentists with limited bitemark experience, and c) 2nd Year Dental Students.

Ten examiners were recruited in each group based upon their ability to meet the established criteria: access to a computer with version Adobe Photoshop® v5.0 or higher, capability to save digital images to either a Zip disk or CD-ROM, and willingness to volunteer their time to the study. A total of thirty examiners received the study materials.

Every individual has a presumed dental uniqueness that is exhibited in the shape, size, position, and anatomical variations found within the teeth and corresponding alveolar structures. Therefore, the selection of the casts used for this study had to reflect a degree of dental uniqueness found in the population. The dental casts were selected to be representative of a typical range of the variations found in the human dentition. Six casts were chosen that exhibited assorted levels of difficulty based upon tooth placement and rotation. The casts were stratified as follows: two casts had near perfect alignments, two casts showed moderate levels of crowding, and two casts had buccal-lingual inversions.

Digital images of the casts were obtained by scanning to high-quality JPEG format using the method described by Johansen (2). An ABFO No. 2 scale was included in each image. The final images were confirmed to be life-size and were saved in a Photoshop Document (PSD) file format at a resolution of 150 pixels/in.

In order to evaluate intra-examiner reliability, it was necessary that some of the casts would need to be examined more than once. To accomplish this, all cast images were duplicated and altered at the occlusal surfaces of the posterior teeth to mislead the examiners into thinking the casts were different.

To control for the potential for the examiners to improve their skills with each successive overlay, the order the casts were to be completed was randomized. Each examiner was placed into one of ten participant groups and the cast order was randomized within these groups.

Each of the six original casts was duplicated and modified resulting in twelve casts images that were distributed to the examiners. The casts were labeled Cases 1 through 12 dependent upon which randomization group the examiner was placed in. Each examiner would create an overlay for each case twice allowing for calculations of intra-examiner reliability without the need for a washout period to elapse between examinations. A questionnaire was developed to gauge experience levels for each examiner. The questionnaire focused on two areas: computer and bitemark experience.

Each examiner received a package including the following: Instructions, a Questionnaire, a CD-ROM labeled "Study CD" containing the twelve cast images labeled Cases 1–12, a CD-ROM or Zip Disk labeled Blank CD or "Blank Disk" to submit their overlays, and a pre-addressed envelope for the return of the materials. It was anticipated that the time commitment for each participant would be approximately two to four hours, and this was determined to be the maximum, reasonable time requested of the participants.

Techniques

Technique A is the well-known, peer-reviewed method (3,4) routinely used in North America. The main feature in Technique A is the use of the Magic Wand Tool. The Magic Wand cursor is placed over the biting edge of each tooth and the area is selected. The area selected will include all adjacent pixel tones of similar values based upon the tolerance setting of the Magic Wand Tool. This will create an outline of the perimeters of the biting surface of each tooth.

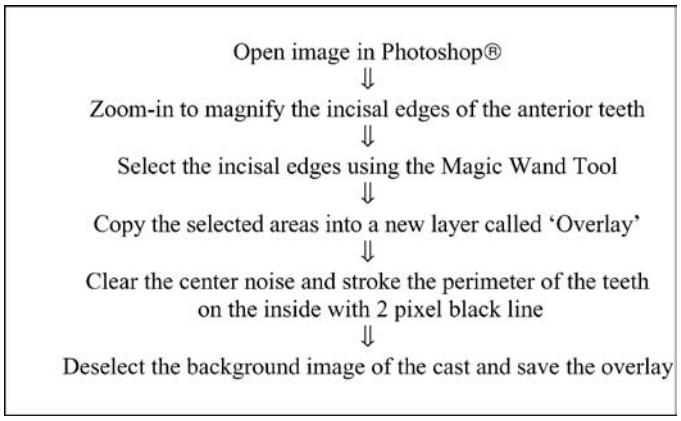

FIG. 1—*Technique A overlay production steps.*

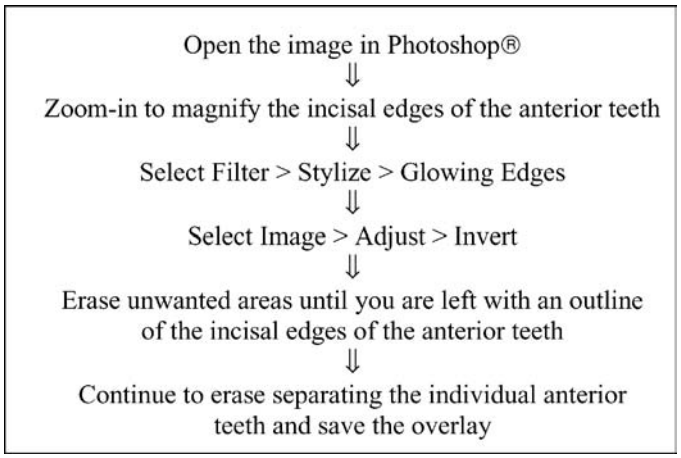

FIG. 2—*Technique B overlays production steps.*

The perimeter is then defined by stroking a solid black line around the selected area to produce the overlay. Figure 1 summarizes the process involved.

Technique B is used predominately in Europe and is based on previously published methods (7,8) that focus on different contrast levels found in the cast image. The main feature in Technique B is the use of the Glowing Edges filter. This filter independently selects and highlights the edges of the cast and teeth. The contrast of the image is then inverted and the unwanted areas are erased until only an outline of the incisal edges of the anterior teeth remain. Figure 2 summarizes the process involved.

After the examiners returned the overlays, a follow-up questionnaire was used to collect additional data on technique preferences, time involved and magnification levels.

Measurements of Area and Position

The renderings of the biting edges of the teeth recorded on the overlays by the examiners were converted to black using Adobe Photoshop®. The final product was saved as an uncompressed TIFF image. Figures 3 and 4 illustrate one of the examiner's original hollow-volume overlays and the converted solid-volume overlays for each technique.

If two teeth were found adjoined, a one-pixel wide break was created using the eraser tool at the normal anatomical separation between the teeth. This was necessary to differentiate each tooth.

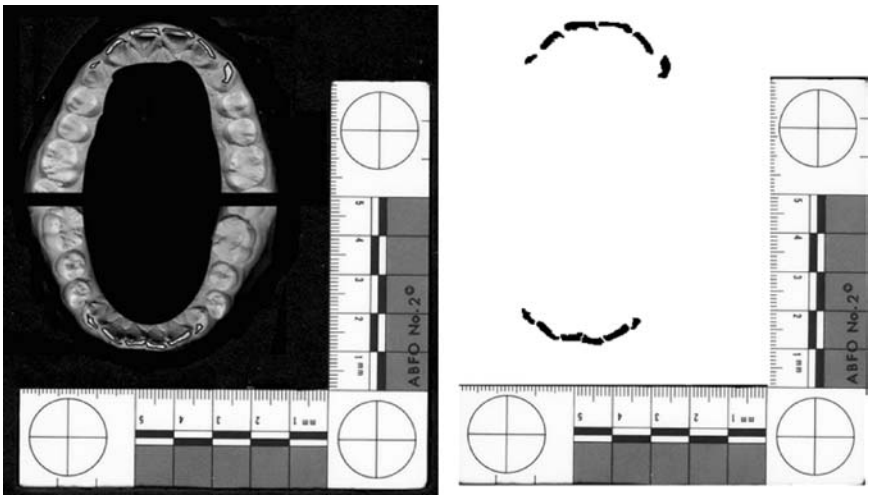

FIG. 3—*Original and converted overlay example for Technique A.*

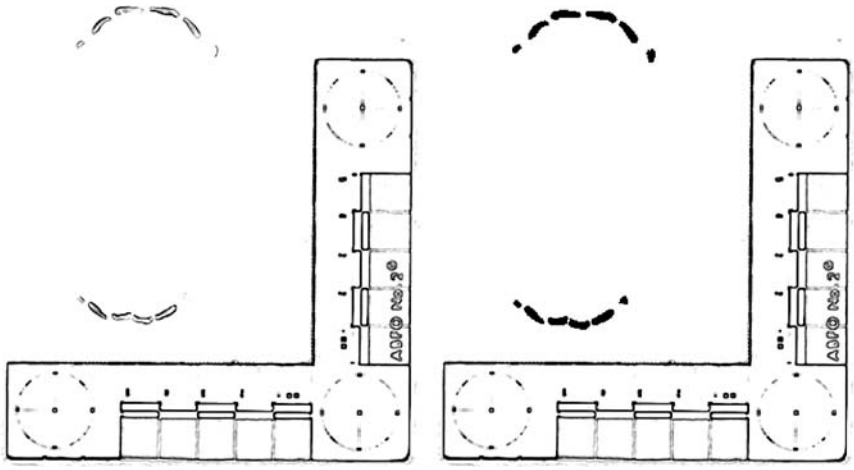

FIG. 4—*Original and converted overlay example for Technique B.*

Each converted overlay was opened in Scion Image® and the following preferences were set: a) area measurements in millimeters, b) x–y centroid position, and c) activation of the wand auto-measure tool. Measurements of area, x-coordinate, and y-coordinate were collected for each tooth on each converted overlay. This resulted in 36 data points for each case. A total of 432 data points were collected for each examiner for each technique.

Assessment Methods

A suitable model to assess the reliability for the parameters of the study was developed.

The three outcome measures that have been obtained for each tooth are: 1) area in square millimeters, 2) x-coordinate of the centroid position, and 3) y-coordinate of the centroid position. All of the outcome measures were analyzed independently. Two assessment methods were used for evaluation: an analysis of variance and calculation of reliability coefficients.

A SAS PROC GLM program was used to perform the ANOVAs. The observed mean squares from the ANOVAs were put into a Quattro-Pro spreadsheet to calculate the reliability coefficients. The results were then compared.

Results

Reliability Assessment

A reliability assessment method was required to analyze the degree of agreement between the examiners. The calculation of a reliability coefficient was determined for each outcome measure, for each group of observers, and for each technique under investigation. Additionally, the inter- and intra-examiner assessment was evaluated. A coefficient value will range between zero and one, whereas a value close to one indicates excellent reliability and a value close to zero indicates poor reliability. Twenty-four examiners returned the completed overlays and were included in the final analysis. The results for reliability are displayed in Table 1 for each examiner group and a cumulative total.

In the assessment of the results for area measurements, the overall inter-examiner reliability coefficient found for Technique A was 0.327 and for Technique B was 0.437. The intra-examiner reliability coefficients results for Technique A and B were slightly better with values of 0.527 and 0.550, respectively. Each examiner group exhibited a similar trend with ranges that fell between 0.260 to 0.651. These values indicate a less than adequate measure of reliability.

The assessment of the results for the positional measurements yielded significantly higher means. The overall inter-examiner

TABLE 1—*Comparison of the inter- and intra-examiner reliability coefficients for all examiner groups and a cumulative analysis for each variable analyzed.*

	Inter-examiner		Intra-examiner	
	Tech A	Tech B	Tech A	Tech B
Diplomates Area	0.421	0.510	0.651	0.615
X-position	0.952	0.973	0.953	0.974
Y-position	0.997	0.998	0.997	0.999
Dentists Area	0.260	0.349	0.478	0.489
X-position	0.942	0.949	0.943	0.949
Y-position	0.996	0.990	0.996	0.990
Students Area	0.423	0.522	0.503	0.584
X-position	0.940	0.968	0.941	0.968
Y-position	0.996	0.992	0.996	0.992
Cumulative Group Area	0.327	0.437	0.527	0.550
X-position	0.946	0.964	0.947	0.964
Y-position	0.997	0.993	0.997	0.993

TABLE 2—*Area ANOVA Results for Technique A and B.*

		Area			
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Technique A					
Experience Level	2	1160.566	580.283	95.52	<0.0001
Examiner	21	21918.196	1043.724	171.81	<0.0001
Cast	5	3074.495	614.899	101.22	<0.0001
Tooth	66	20109.208	304.685	50.16	<0.0001
Technique B					
Experience Level	2	5588.747	2794.373	232.08	<0.0001
Examiner	21	21074.677	1003.556	83.35	<0.0001
Cast	5	3737.388	747.478	62.08	<0.0001
Tooth	66	47167.056	714.652	59.35	<0.0001

reliability coefficients calculated for Technique A for both the x- and y-position were 0.946 and 0.997. Similarly, the values for Technique B were 0.964 and 0.993. The results were as impressive in the evaluation of intra-examiner reliability. The reliability coefficients were calculated to be 0.947 and 0.997 for Technique A and 0.964 and 0.993 for Technique B. These results confirmed a high degree of reliability for both techniques in the assessment of the positional data.

ANOVA Assessment

An analysis of variance is commonly employed when comparing the effects of different factors between different groups of examiners. The analysis was required to assess the effect of the following factors: forensic experience, examiners, casts, teeth, and random error. The ANOVA analysis for the area, x-position, and y-position are displayed in Tables 2 through 4.

The analysis of variance calculations for the area measurements indicate that all four variables analyzed for their contribution to the effect size, (experience level, examiner, cast, and tooth) resulted in a p value of <0.0001 for both techniques. This suggests that each variable had a significant effect on the differences in the means, thus eliminating the chance the differences were caused by random error alone.

TABLE 3—*X-Position ANOVA Results for Technique A and B.*

		X-Position			
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Technique A					
Experience	2	28.0367	14.018	1.53	0.2157
Examiner	21	853.883	40.661	4.45	<0.0001
Cast	5	11589.475	2317.895	253.74	<0.0001
Tooth	66	519046.890	7864.347	860.91	<0.0001
Technique B					
Experience	2	15.284	7.642	1.41	0.2453
Examiner	21	391.523	18.6439	3.43	<0.0001
Cast	5	3108.507	621.7014	114.36	<0.0001
Tooth	66	466575.485	7069.3255	1300.41	<0.0001

TABLE 4—*Y-Position ANOVA Results for Technique A and B.*

		Y-Position			
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Technique A					
Experience	2	17.077	8.538	1.69	0.1845
Examiner	21	185.302	8.824	1.75	0.0187
Cast	5	1183.102	236.620	46.85	<0.0001
Tooth	66	4627875.096	70119.320	13884.8	<0.0001
Technique B					
Experience	2	33.978	16.989	1.80	0.1657
Examiner	21	903.165	43.008	4.55	<0.0001
Cast	5	258045.290	51609.058	5463.87	<0.0001
Tooth	66	4518747.050	68465.864	7248.51	<0.0001

The analysis of variance results for Technique A indicated that for the x-position, the examiner, cast, and tooth all had a p value <0.0001, translating to a significant effect on the differences in the means. However, the experience factor had a larger p value of 0.2157. This value implies that experience level was responsible for less of a significant effect in the means than the other factors. This same trend was noted for Technique B with the experience factor having a p value of 0.2453.

Comparable results were shown with the evaluation of the y-position. The analysis of variance results for Technique A for the y-position showed that the cast and tooth had significant effects on the differences in the means with both factors having a p value of <0.0001. The effects of the experience level and observer had a lesser effect with p values of 0.1845 and 0.0187, respectively. These results support the notion that the forensic experience level of the examiners has less of an effect on the differences than the cast and the teeth. Similarly, Technique B showed the experience level to have less of an effect than the other factors with a p value of 0.1657.

Discussion

Examiners

In a previous study (5), three examiner groups were also chosen to represent differing levels of bitemark experience. The study's conclusions showed no significant differences between the three groups suggesting that experience and training in bitemark analysis have no affect on the success of using overlays to identify biters

correctly. The current study focuses on having the examiners produce the overlays, rather than the use of an overlay to analyze a bitemark.

Recruitment of examiners to participate in this study was a challenge due to the time commitment. After having committed to the project, a few examiners decided they no longer had the available time and withdrew from the study. Furthermore, some examiners revealed, after the fact, that they lacked the version of Adobe Photoshop® necessary to complete the project and were forced to withdraw. The overall response was an 80% return rate. The end result was eight Diplomates, eight forensic dentists, and eight students. In future studies, it would be beneficial to collect data and information in a workshop or symposium. A controlled environment, standardization of computer software and hardware, and immediate acquisition of the results would eliminate some of the obstacles that were encountered in the collection of this data.

Cast Selection

There were different degrees of difficulty randomly shown in each cast. To increase compliance, both inter- and intra-examiner studies were completed simultaneously in one session. Some examiners commented that they noticed a few cast images were duplicates but appreciated being able to complete the study in one setting. The results indicate an accurate assessment of intra-examiner results can be obtained without the need for a wash-out period. The analysis of variance showed that all casts and individual teeth contribute to significant variations of the means. Therefore, the specific degree of difficulty associated with each cast had no significant effect on the results.

Area Measurements

The reliability coefficients obtained for the area measurements in Technique A were considerably lower than expected. The unfavorable results can be explained with an examination of a trend seen in some of the overlays submitted. For some novice examiners, and some experienced examiners with limited computer practice, there was a tendency to produce and overlay by simply single clicking the Magic Wand Tool on one area of the biting surface of each tooth. If the selection was in an area of different contrast levels of pixels than the majority of the remaining biting surface, such as a chip or worn edge of a tooth, the Magic Wand tool would limit the highlighted area to that specific region. Instances of single clicking were found in all levels of examiners suggesting that this was not a result of lack of knowledge regarding bitemarks, but rather lack of experience with the use of Adobe Photoshop®. In addition, the instructions for Technique A should be clearer as to the likelihood of having to select multiple clicks for each individual tooth. Having the examiners print out the overlays to use in an actual bitemark comparison could eliminate the occurrence of this type of mishap by allowing the examiners to directly evaluate the applicability of the overlay they produced.

The reliability coefficients calculated for Technique B were slightly greater for all examiner groups. Again, these values fell beneath an acceptable reliability level.

While Technique A involved selecting areas to include in an overlay, Technique B focused on removing unwanted areas to result in an overlay. An overwhelming majority of participants commented on the tedious task of erasing the unnecessary portions. The lack of control of the mouse, inability to differentiate the tooth surface from the cast surface in a contrast-inverted image, and the fickle nature of the Eraser tool were criticism of this technique.

The attention to detail, or lack there of, to define the individual perimeters of the teeth was evident in a few overlays. While some examiners clearly defined each individual tooth, others simply followed the curvature of the teeth to produce a general outline of the arch. This factor may have contributed to the extreme variability found within the area of each tooth. By not erasing the excess "noise" around the perimeters of each tooth, a substantial amount of additional area becomes included in the overlay.

Position Measurements

The reliability coefficients calculated for all examiner groups were exceptionally high with near perfect agreement in some cases. These results help demonstrate the objective advantage of using computers to determine positional data. Of additional interest is the fact that there were no significant differences in the values obtained for each examiner group seen in the ANOVA results. This suggests that previous bitemark experience is not a significant factor in the production of an overlay.

The reliability coefficients for Technique B were equally remarkable. The positional reproducibility by all examiner groups strengthens the notion that computer-generated bitemark overlays can be consistently and objectively produced.

Individual Examiner Preferences

The overwhelming majority of participants preferred Technique A. The most often cited comment was the ability to accurately and clearly interpret the cast image. With a true representation of the cast visible, examiners said they had more control in selecting the biting surfaces based on their interpretation of the anatomy of the teeth. The average time required to complete an overlay using Technique A was faster than using Technique B.

Criticisms of Technique A centered on the fastidious nature of the Magic Wand Tool. Novices in the production of computer-generated overlays remarked on the difficulties experienced in selecting the biting edges. This problem can be resolved by adjusting the Tolerance setting of the Magic Wand Tool. Examiners unfamiliar with Adobe Photoshop® may not be aware of the significance of adjusting the tolerance level of the Magic Wand Tool or even have the knowledge to be able to modify the settings. Clarification of this step should be included in any further studies.

Interestingly, most participants found Technique B to be more subjective than Technique A. This response may be in part to the differences in each examiner to interpret a negative image. The greatest complaint was the loss of the true image of the cast after the contrast inversions. The few examiners who preferred Technique B to Technique A felt that the Eraser Tool was easier for them to control than the Magic Wand Tool.

Technique A is more commonly used in North America and because all the examiners reside in North America, this may be a contributing factor. However, the partiality towards Technique A was seen in the group of dental students as well, who would have no previous familiarity with any established techniques.

Conclusions

The reliability assessment of Technique A and Technique B produced similar results. Both techniques scored low for area measurements and extremely high for positional measurements. The study concluded that both techniques are reliable methods to produce bitemark overlays to assess tooth position.

While investigators have studied the reliability of the use and production of bitemark overlays, the next step is to test the external validity of bitemark overlays. This would include having examiners both produce and utilize their original overlays in a simulated bitemark comparison. Additional studies comparing overlays should assess the differences in tooth rotation and evaluating the differences seen in the area and positional measurements produced by the same cast with different degrees of penetration into the test medium. As with any study that involves outside examiners to follow a specified technique, the importance of the training tools, such as directions and instructions, cannot be over-emphasized. Future overlay studies should include a quantitative test of the instructions prior to assessing the technique.

Further, studies regarding the effects of different scanning procedures could provide insight as to the best way to obtain cast images. Examining different variables such as scanning resolutions, placement of the casts, and different dental stone materials may aid in standardizing optimal procedures.

Finally, as computer software programs continue to be modified and upgraded from previous versions, the changes in the features should continue to be examined for potential benefits. Whether the software may enhance the bitemark photograph or assist in the examination of the suspect's cast, the use of computers will continue to advance bitemark analysis.

Acknowledgments

The authors gratefully acknowledge Drs Douglas Waterfield, Donald Brunette, David Lowe, Pierre Dow, and Alan Donald for their research advice and assistance, as well as the study's participants. Dr. McNamee is supported by a research grant from the American Board of Forensic Odontology, the American Society of Forensic Odontology, the University of British Columbia Faculty of Dentistry Research Grant, and the Bureau of Legal Dentistry.

References

1. Bowers CM. Jurisprudence issues in forensic odontology. Dent Clin North Amer 2001;45(2):399–415.
2. Sweet DJ, Bowers CM. Accuracy of bite mark overlays: a comparison of five common methods to produce exemplars from a suspect's dentition. J Forensic Sci 1998;43(2):362–7. [PubMed]
3. Johansen RJ, Bowers CM. Digital analysis of bitemark evidence. Santa Barbara, CA: Forensic Imaging Services, 2000.
4. Sweet D, Parhar M, Wood RE. Computer-based production of bite mark comparison overlays. J Forensic Sci 1998;43(5):1050–5. [PubMed]
5. Pretty IA, Sweet D. Digital bite mark overlays-an analysis of effectiveness. J Forensic Sci 2001;46(6):1385–91. [PubMed]
6. Robinson E, Wentzel J. Toneline bite mark photography. J Forensic Sci 1992;37(1):195–207. [PubMed]
7. Naru AS, Dykes E. The use of a digital imaging technique to aid bite mark analysis. Sci Justice 1996;36(1):47–50. [PubMed]
8. Naru AS, Dykes E. Digital image cross-correlation technique for bite mark investigations. Sci Justice 1997;37(4):251–8. [PubMed]

Additional information and reprint requests:
David Sweet, D.M.D., Ph.D.
Bureau of Legal Dentistry
University of British Columbia
146-2355 East Mall
Vancouver, BC, Canada V6T 1Z4
E-mail: dsweet@interchange.ubc.ca