J Forensic Sci, January 2007, Vol. 52, No. 1
doi:10.1111/j.1556-4029.2006.00321.x
Available online at: www.blackwell-synergy.com

Stella Martin-de las Heras,[1] M.D., Ph.D., B.D.S.; Aurora Valenzuela,[1] M.D., Ph.D., B.D.S.; A. Javier Valverde,[1] B.D.S.; Juan Carlos Torres,[2] Phys.D., Ph.D.; and Juan D. Luna-del-Castillo,[3] Ph.D.

Effectiveness of Comparison Overlays Generated with DentalPrint© Software in Bite Mark Analysis*

ABSTRACT: Validation studies of the new DentalPrint© software were carried out with experimental bite marks in pigskin. The bite marks were digitally photographed according to the ABFO guidelines for evidence collection. Dental casts used in the experiment were scanned in 3D and 2D, and comparison overlays were generated using DentalPrint© and Adobe® Photoshop® software, respectively. Digitized photographs of the experimental bite marks and the biting edges obtained in the overlays were compared by two different examiners to analyze the impact of training and experience with the two methods. Receiver operating characteristic (ROC) analysis, sensitivity, specificity, and 95% confidence intervals for each cutoff point were calculated. The expert examiner using DentalPrint© obtained the best results, with an area under the ROC curve of 0.76 (SE = 0.057; CI at 95% = 0.652–0.876). Fairly high specificity values were found for DentalPrint©, and the best results were obtained for the cutoff value that discriminated between the examiner's response "biter" and the rest of the possible (specificity 97.9%, CI at 95% = 93.2%–99.6%). Therefore, the results presented here indicate that DentalPrint© is a useful, accurate tool for forensic purposes, although further research on the comparison process is needed to enhance the validity of bite mark analysis.

KEYWORDS: forensic science, forensic odontology, bite marks, 3D images, DentalPrint©, accuracy, effectiveness

Bite mark analysis, from a forensic perspective, is founded on two premises. The first is that each individual's dentition is presumed to be unique, and the second is the assumption that uniqueness of the dentition is accurately recorded in the injured skin or in objects. As a result, bite mark evidence has been almost universally accepted in courts of law. However, criticism of this evidence as a reliable scientific tool in court proceedings has been voiced because of the subjective nature of the comparative analyses (1–4) and the insufficient published work that defines the statistical parameters of the procedures that form the basis of scientific conclusions (5).

To achieve high professional standards in bite mark analysis, the American Board of Forensic Odontology (ABFO) has adopted guidelines for the analysis procedure and the terminology to express conclusions (6). In accordance with development of these guidelines, the ABFO also promotes the continuing education of forensic odontologists.

The procedures for comparing bite marks to the dentition of possible assailants are well established. Most common methods of analysis produce life-sized comparison overlays from the suspect's teeth to detect similarities or differences with the bite mark (7–15). Some authors have improved the methods for comparative analysis by using computer-based techniques to produce bite mark comparison overlays (12–14). In these methods, dental casts are scanned with a two-dimensional (2D) scanner and images are imported to Adobe® Photoshop® software (Mountain View, CA). The biting edges of the suspect's teeth are selected based on similarities between adjacent pixel values. Tests of this procedure have become a first step in establishing the scientific basis of this technique in forensic dentistry (16).

We recently published a new method, DentalPrint© (2004, Department of Forensic Medicine and Forensic Odontology, University of Granada, Granada, Spain), to generate comparison overlays from 3D images of the suspect's dental casts (17). The procedure has the advantages of being entirely automatic, thus avoiding observer bias, making it impossible for third parties to manipulate or alter the 3D images, and generating different comparison overlays depending on the pressure of the bite or the distortion caused by victim–biter interaction.

In spite of these advantages, forensic odontologists must ensure that the techniques they use are supported by scientific evidence to promote justice. Therefore, the purpose of this study was to determine values of intra- and interexaminer reliability, sensitivity, specificity, and validity for the new DentalPrint© software. Comparison studies with another well-established system of overlay generation, i.e., Adobe® Photoshop® software, are also reported here. The impact of the examiners' training and experience was also measured.

Materials and Methods

Experimental Bite Marks

Seventeen dental casts exhibiting variations in the presence, status, and arrangement of their upper or lower anterior teeth were selected for this study. Each set of models was given a different code number and was then used to produce experimental bite marks in pigskin as an accurate analog of human skin (18,19).

Freshly slaughtered, eviscerated piglets 4–5 weeks old and weighing 5–6 kg each were obtained from a local abattoir. Postmortem

[1]Department of Forensic Medicine and Forensic Odontology, School of Medicine University of Granada, 18071 Granada, Spain.
[2]Department of Software Engineering, University of Granada, 18071 Granada, Spain.
[3]Department of Biostatistics, University of Granada, 18071 Granada, Spain.

*Preliminary results were presented at the 53th Meeting of the American Academy of Forensic Sciences, February 2001, Seattle, WA.

Received 21 Jan. 2006; and in revised form 7 May 2006, and 9 July 2006; accepted 5 Aug. 2006; published 8 Dec. 2006.

interval ranged from 3 to 10 h. Anatomical locations with minimal skin curvatures and distortion were selected in each piglet. The dental casts from each selected biter were hinged and clamped to the skin for 15 m to create a bite mark.

The bite mark was digitally photographed following the ABFO guidelines for evidence collection (6) with the ABFO No. 2 scale in place. The photographs were stored in JPEG format in a personal computer, and the best reproduction of each bite mark was selected.

Production of Overlays

The production of comparison overlay with DentalPrint© (2004, Department of Forensic Medicine and Forensic Odontology, University of Granada, Granada, Spain) was described in detail in a previous publication (17). Briefly, the procedure was as follows: dental casts were scanned with a 3D contact-type scanner (Picza 3D Scanner® model PIX-3, Roland DG Corp., Shizuoka, Japan). The 3D images were imported to the DentalPrint© software and processed. Comparison overlays from 3D images of the dental casts were obtained in three steps. First, the teeth involved in the bite mark were identified. In the second step, a contact plane was created from the three highest points detected in areas defined in the 3D images of the dental casts. Finally, biting edges were obtained with DentalPrint©, which allows the contact plane to extend deep into the teeth (Fig. 1). The perimeter of the suspect's biting edges can be printed on transparent acetate film or converted into a bmp file. Comparison overlays were also generated with Adobe® Photoshop® software as described previously (12–14).

Comparison Process

The digitized photographs of the experimental bite marks and the biting edges obtained with DentalPrint© software were compared according to the Johansen and Bowers method (14). In this study, one comparison overlay generated from a unique set of dental models (upper and lower) was compared with one bite mark photograph. Each set of models and the bite mark photograph were given a code number. The comparison procedure was performed in accordance with the biostatistician's expert opinion as follows: the observers—one expert and one nonexpert—were given a template (Table 1) where the sequence of comparisons was established, with an appropriate number of comparisons (128 comparisons were made in total) and an adequate number of positive identifications of biters ($n = 32$). All comparisons were made in a blind manner: the observers were unaware of which overlays, or how many of them, were likely to yield a positive identification of the biter.

The ABFO scoring system (6) was used, and examiners' conclusions were recorded as "nonbiter," "probable biter," "possible biter," or "biter." A bite mark photograph from our experimental study, a detail of the overlay generation using DentalPrint© software, and the comparison process made by an expert examiner are shown in Fig. 2.

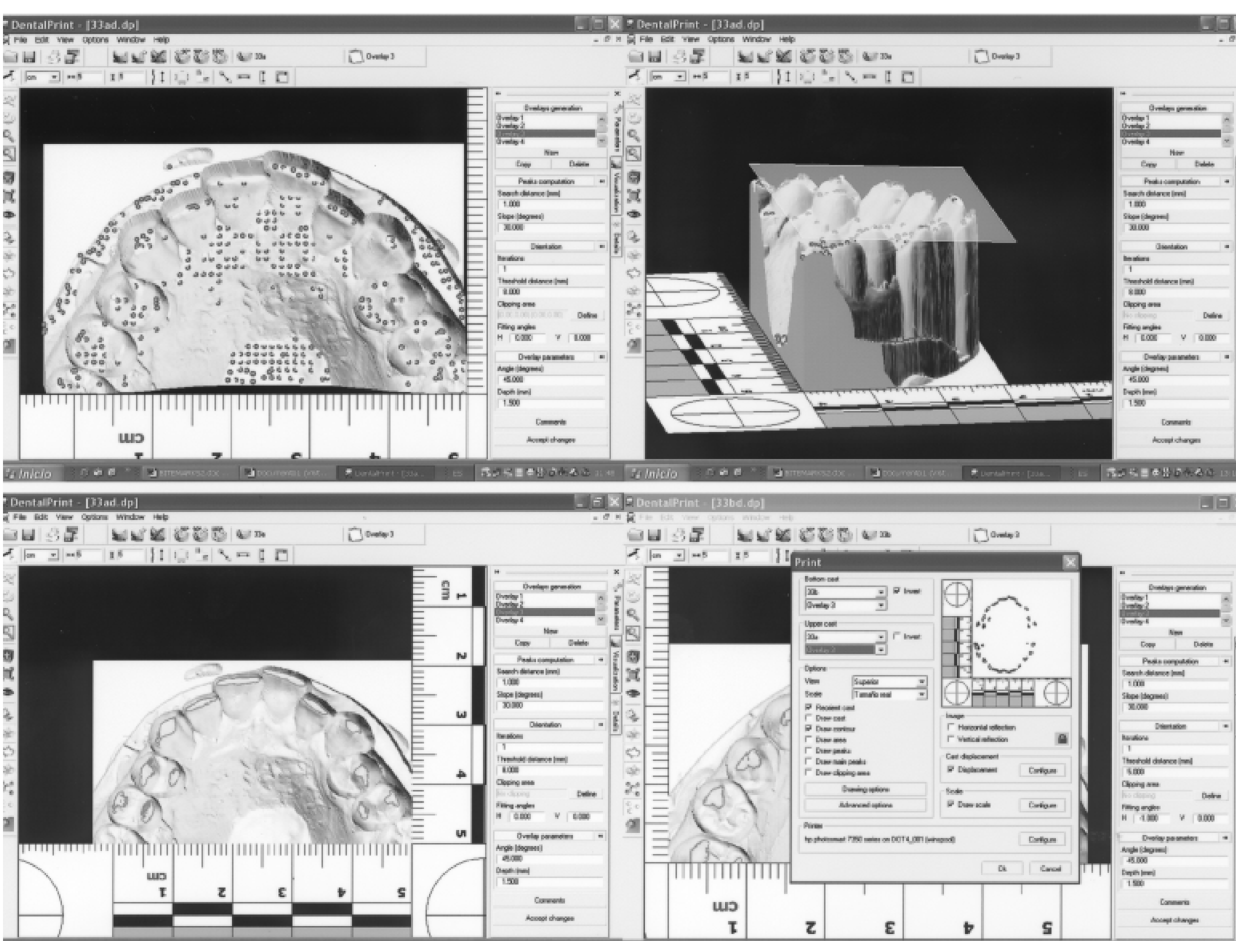

FIG. 1—*Comparison overlay generation from 3D images of dental casts with the DentalPrint© software.*

TABLE 1—*Pattern of the comparison procedure.*

Bite Mark Photograph	Comparison Overlays			
7	33	22*	51	37
16	30	41*	38	56
13	26	51*	45	18
12	51	45*	1	37
16	38	40	41*	37
2	26	38*	22	40
10	45	22	33*	17
14	40*	38	17	1
15	17	51	56*	22
11	37*	17	22	18
14	40*	30	56	30
1	38	40	37	1*
10	30	51	33*	40
2	21	37	38*	1
6	21*	33	40	38
4	30	45	17*	37
5	56	18*	21	51
1	37	21	45	1*
8	26*	45	17	22
9	17	30*	37	13
8	26*	38	18	56
3	38	13*	51	18
3	26	18	51	13*
12	51	1	45*	37
9	45	33	37	30*
5	51	21	17	18*
4	17*	37	30	33
7	51	26	45	22*
15	40	56*	13	1
6	13	17	21*	40
11	13	30	56	37*
13	37	51*	22	41

In this study, one bite mark photograph was compared with one comparison overlay produced from a unique set of dental models (upper and lower). The sequence of comparisons was vertical. The observer started by comparing bite mark photograph number 7 with comparison overlay number 33. After finishing the first column of comparisons, the observer continued with comparisons listed in the second column, and so on. The comparison process finalized by comparing photograph number 13 to comparison overlay number 41.

*Positive match between biter and bite mark injury.

To analyze the impact of training and experience, two different examiners were used: a Ph.D. student trained in bite mark analysis, and a general dental practitioner with an interest in forensic dentistry but with no experience in bite mark analysis. In the case shown in Fig. 2, both examiners' conclusions were of a "biter."

The same comparison process was performed with the overlays generated from Adobe® Photoshop® software. The same two observers compared bite marks and overlays after a 3-month delay from when they had tested the DentalPrint© overlays, to avoid bias in their conclusions.

Statistical Analysis

Receiver operating characteristic (ROC) analysis was used to determine the accuracy of the method. The ROC curve combines the concepts of sensitivity and specificity into a single measure of accuracy, defined as the area under the ROC curve. This was therefore considered to reflect the examiner's ability to recognize correctly the dentition that made the bite mark. Sensitivity, also called the true positive fraction (TPF), was defined as the proportion of correct identifications of dentitions that made the bite mark. Specificity was the proportion of correct identifications of the dentition that did not make the bite mark. In ROC analysis, the false-positive fraction (FPF), i.e., the number of incorrect identifications, is the complement of specificity (1-specificity). When the examiner's responses are expressed as the degree to which he or she believes the bite mark was produced by a particular dentition, each degree can be used as a cutoff point to create an array of TPF/FPF pairs. Plotting these pairs with FPF in the *X*-axis and TPF in the *Y*-axis forms the ROC curve. Areas under the ROC curve between 0.5 and 1 indicate a positive relationship between the rating scale and correct identification. Ninety-five percent confidence intervals for accuracy were calculated using the bootstrap technique.

A secondary objective in this study was to determine whether examiner experience had an effect on accuracy. Differences between areas under the ROC curve for each of the two examiners were compared with the χ^2 test, and κ values were calculated to measure agreement between the examiners' conclusions and to control for chance agreement.

Results and Discussion

In bite mark analysis, two different experimental materials can be used to study the effectiveness: real forensic cases or simulated cases. In this study, we decided to use simulated bite marks. The use of real cases has advantages because authenticity is assured; however, one of the criteria for assessing the effectiveness of a particular test is to compare its performance against a suitable gold standard. The use of real case material would require us to use the conclusions of the Court decision or the original examining odontologist as the standard, although the possibility of error can never be ruled out completely. Therefore, the use of real forensic cases to determine the truth is a potential weakness in studies of the effectiveness of bite mark analysis (20). The use of simulated material, however, also has disadvantages. Nonhuman skin and the postmortem bites, as used in this study, do not display any of the vital patterns seen in antemortem bite injuries. However, in simulations, there is no doubt about the identity of the biter.

DentalPrint© software generates comparison overlays and avoids the bias inherent in observer subjectivity, as the entire procedure for generating overlays is automatic (17). As previously reported (17), several experiments were carried out to analyze the reliability and objectivity of the new software. Moreover, excellent intra-class correlation coefficients (ICC = 0.9985) and confidence intervals (CI) at 95% (0.9959–0.997) were obtained for all measurements by the same examiner. Interobserver reproducibility of measurements obtained with DentalPrint© software was also excellent, with an ICC of 0.9999 and CI at 95% of 0.999–1.000 (17).

To determine the accuracy of the DentalPrint© software, we used ROC analysis to graph the reciprocal relationship between sensitivity and specificity calculated from all possible threshold values. Results from the 128 comparisons by an expert examiner who used DentalPrint© to produce comparison overlays are shown in Table 2, and the ROC curve is shown in Fig. 3. The area under the ROC curve is 0.76, a fairly high value according to Swets' scoring system (21). The standard error is 0.05, and the 95% confidence interval is 0.652–0.876. The area under the curve provides an objective parameter of diagnostic accuracy of the test that is far superior to comparing single combinations of sensitivity and specificity. The results presented in this study indicate that bite mark examination with DentalPrint© is an accurate forensic technique and the examiner is able to identify correctly the dentition belonging to a particular bite mark.

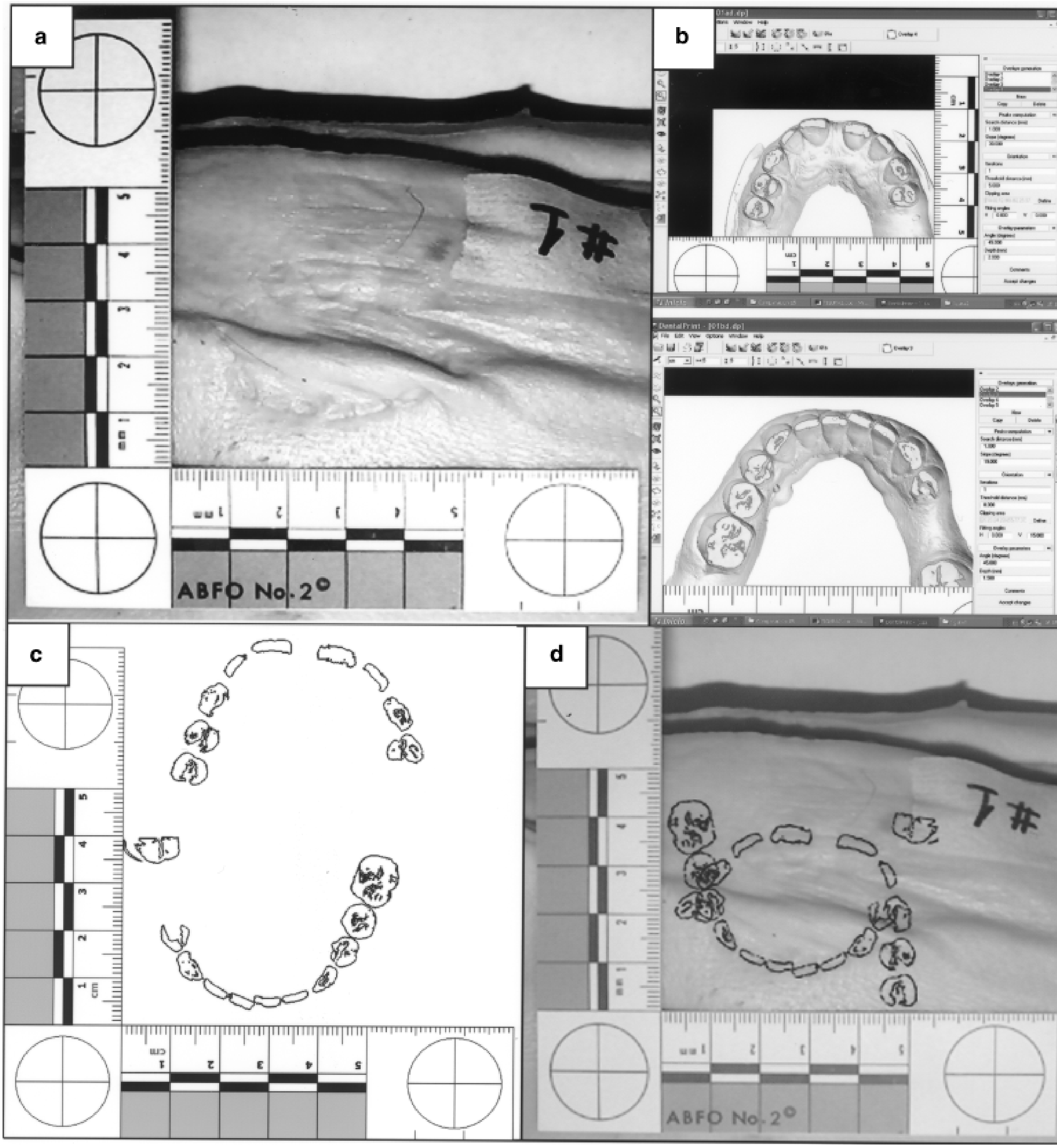

FIG. 2—*Experimental study of bite marks. Example of one bite mark photograph (a); overlay generation from 3D images of a set of dental casts with the DentalPrint© software (b and c); and the final comparison process (d).*

The sensitivity, specificity, and 95% confidence intervals for each specific cutoff point are shown in Table 3. No statistically significant differences in sensitivity or specificity were found between different cutoff points. The modest sensitivity values that we obtained mean that this method should be considered with caution for screening test. The best cutoff point was that which discriminated between the examiner's response "nonbiter" and the rest of the possible, which had a sensitivity of 56.3% and CI at 95% from 38.5% to 72.2%.

However, forensic tests should minimize false positives to avoid an erroneous suggestion of guilt in an innocent person. Therefore, the main aim of forensic tests for bite marks should be to identify correctly dentitions that did not make the bite mark. In

TABLE 2—*Results from the ROC analyses for both types of examiners according to the technique used to produce comparison overlays.*

Examiners	AUC	SE	CI	
Expert (3D*)	0.764†	0.057	0.652	0.876
Nonexpert (3D*)	0.642†	0.062	0.521	0.763
Expert (2D‡)	0.726§	0.059	0.610	0.841
Nonexpert (2D‡)	0.598§	0.062	0.477	0.720

*Comparison overlays generated from DentalPrint© software.
†Significant differences between observers ($\chi^2_{exp} = 4.09$; 1 df, $p \leq 0.05$).
‡Comparison overlays generated from Adobe® Photoshop® software.
§Significant differences between observers ($\chi^2_{exp} = 4.81$; 1 df, $p \leq 0.05$).
AUC, area under the ROC curve; SE, standard error; CI, confidence intervals at 95%.

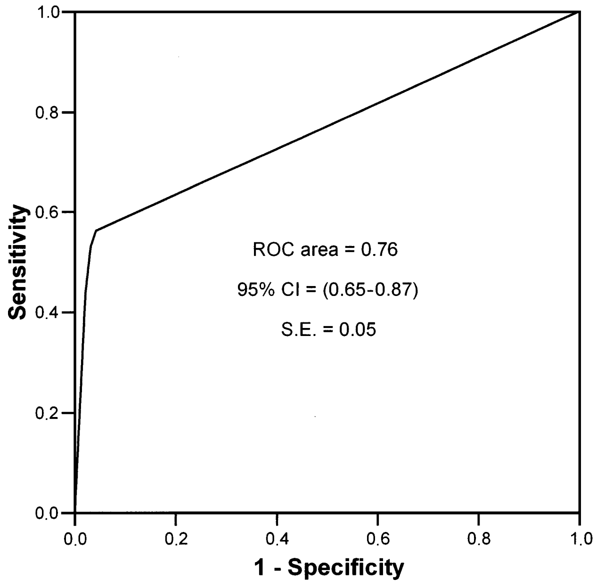

FIG. 3—*Receiver operating characteristic (ROC) curve generated from the expert examiner's conclusions with DentalPrint© software to produce comparison overlays.*

this study, fairly high specificity values were found, with the best result for the cutoff point that discriminated between the examiner's response "biter" and the rest of the possible. Specificity was 97.9%, with a CI at 95% of 93.2–99.6%.

As the experimental methods used were not identical to previously assayed methods, our results should be compared with earlier analyses of bite mark with due caution. One earlier study by Pretty and Sweet (16) analyzed reports by 10 ABFO diplomates (with the highest level of training and experience), who examined 10 simulated bite mark cases using Adobe® Photoshop® software to generate comparison overlays. In that study, the area under the ROC curve was 0.80, with a standard error of 1.18. Therefore, although our ROC values were slightly lower than those obtained by Pretty and Sweet, it should be recalled that the two approaches are not comparable. Future studies are planned with larger numbers of experts with different levels of training and experience in bite mark analysis to document the interobserver reliability with greater precision.

A practicing dentist who was not an expert in bite mark analysis but who had an interest in forensic dentistry analyzed the same 128 cases as the expert examiner (Table 2). The area under the curve was 0.642, standard error was 0.06, and CI at 95% was 0.521–0.763. Although the experimental designs and techniques differed between the present study and theirs, these results are slightly lower than the values obtained by Pretty and Sweet (16) for examiners with the same level of training.

TABLE 4—*Agreement between expert and non-expert examiners for the different cutoff points using both methods to produce comparison overlays.*

Cutoff	κ	κ_{inf}	κ_{sup}
3D†			
0–1, 2, 3	0.805	0.731	0.867
0, 1–2, 3	0.859	0.788	0.915
0, 1, 2–3	0.891	0.828	0.936
2D‡			
0–1, 2, 3	0.805	0.731	0.867
0, 1–2, 3	0.875	0.809	0.926
0, 1, 2–3	0.967	0.926	0.989

†Comparison overlays generated from DentalPrint© software.
‡Comparison overlays generated from Adobe® Photoshop® software.
0 = nonbiter; 1 = probable biter; 2 = possible biter; 3 = biter.

To compare the two methods for bite mark analysis (Adobe® Photoshop® software vs. DentalPrint© software), the same two observers, one expert and one nonexpert, analyzed the same 128 bite marks with overlays produced with Adobe® Photoshop® software (Table 2). Regardless of the observer's experience, the best results were obtained when DentalPrint© software was used to produce the comparison overlays.

To study whether training influenced bite mark analysis, areas under the ROC curve were compared with the χ^2 test. Significant differences were found between the expert and nonexpert examiners with both DentalPrint© ($\chi^2_{exp} = 4.09$; 1 df, $p \leq 0.05$) and Adobe® Photoshop® software ($\chi^2_{exp} = 4.81$; 1 df, $p \leq 0.05$; (Table 2). κ values were calculated to measure the agreement between expert and nonexpert observers for the different cutoff points and for the two methods (Table 4). The significant differences between areas under the ROC showed that observer's experience was relevant in the accuracy of bite mark analysis. Agreement between examiners as reflected by the κ index and classified according to the six-category system proposed by Landis and Koch (22) was "perfect" or "substantial." Therefore, in this study training and experience affected the success of overlays in correctly identifying the biter; however, Pretty and Sweet (16) found no significant differences between different groups of qualified examiners.

A critical question is whether this and other methods to analyze bite marks are accurate enough for identification purposes. A diagnostic test is considered "highly accurate" when the area under ROC is 0.9, "useful for some purposes" at 0.7–0.9, and "poor" at 0.5–0.7 (21). Although from the results presented in this paper it might seem that bite mark analysis by the comparison of overlays fell short of excellence, when the comparisons are performed by an expert examiner, significantly better results (areas under ROC higher that 0.7) were found using either of the methods to generate overlays; therefore, the methods can be considered useful for some purposes.

TABLE 3—*Sensitivity and specificity values for each specific cutoff points obtained for the expert examiner using DentalPrint© software.*

Cutoff	Sensitivity (%)	CI: $Sensitivity_{inf}$ (%)	CI: $Sensitivity_{sup}$ (%)	Specificity (%)	CI: $Specificity_{inf}$ (%)	CI: $Specificity_{sup}$ (%)
0–1, 2, 3	56.3	38.5	72.2	95.8	90.2	98.6
0, 1–2, 3	53.1	34.7	70.5	96.9	91.7	99.9
0, 1, 2 –3	43.8	27.8	61.5	97.9	93.2	99.6

0 = nonbiter; 1 = probable biter; 2 = possible biter; 3 = biter; CI = 95% confidence intervals; $sensitivity_{inf}$ and $specificity_{inf}$, minimum values of sensitivity and specificity; $sensitivity_{sup}$ and $specificity_{sup}$, maximum values of sensitivity and specificity.

However, the software used to generate comparison overlays is not the only factor that needs to be examined in experimental studies designed to analyze the effectiveness of bite mark analysis. Although in experimental studies most of the bite marks created are artificially clear and show less distortion than under actual case work, statistical studies of our results from the final comparison process used complicated cases with unclear photographs or significant distortion.

Moreover, because other elements of the comparison process will also influence the conclusions reached by different examiners, and in order to arrive at more robust conclusions regarding the overall effectiveness of these methods for their application in true forensic cases, further research in the comparison process is needed to refine computer-based methods of image processing for use in bite mark analysis.

Acknowledgments

We thank Dr. David Sweet (University of British Columbia, Vancouver, Canada) for his scientific advice and encouragement, and K. Shashok for editorial help with the manuscript.

This work was supported by a grant from the Research and Education Department, Junta de Andalucia, Spain.

References

1. Rothwell BR. Bite marks in forensic odontology: fact or fiction? In: Worthington P, Evans JR, editors. Oral and maxillofacial surgery. Philadelphia: WB Saunders, 1994:588–600.
2. Zarkowski P. Bite mark evidence: its worth in the eye of the expert. J Law Ethics Dent 1988;1(1):47–57.
3. Pretty IA, Sweet D. A comprehensive examination of bitemark evidence in the American legal system. In: Proceedings of the 52th annual meeting of the American academy of forensic sciences; Feb 21–26, 2000, Reno, NV. Colorado Springs, CO: American Academy of Forensic Sciences, 2000:146.
4. Hale A. The admissibility of bitemark evidence. South Calif Law R 1978;51(3):309–34.
5. Sweet D, Pretty IA. A look at forensic dentistry—part 2: teeth as weapons of violence—identification of bitemark perpetrators. Br Dent J 2001;190(8):415–8.
6. American Board of Forensic Odontology. ABFO bite mark methodology guidelines. In: Bowers CM, Bell GL, editors. Manual of forensic odontology. Montpelier, VT: American Society of Forensic Odontology, 1995:334–7.
7. Cameron J, Sims BG. Bite marks. In: Cameron J, Sims BG, editors. Forensic dentistry. Edinburgh: Churchill Livingstone, 1974:129–45.
8. West M, Friar J. The use of videotape as an aid in bite mark analysis: a case report. J Forensic Sci 1987;34:85–95.
9. Farrel WL. Computerized axial tomography as an aid in bite mark analysis: a case report. J Forensic Sci 1987;32:266–72.
10. Dailey JC. A practical technique for the fabrication of transparent bite mark overlays. J Forensic Sci 1991;36:565–70.
11. Naru AS, Dykes E. The use of a digital imaging technique to aid bite mark analysis. Sci Justice 1996;36(1):47–50.
12. Sweet D, Parhar M, Wood RE. Computer-based production of bite mark comparison overlays. J Forensic Sci 1998;43(5):1050–5.
13. 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.
14. Johansen RJ, Bowers CM. Digital analysis of bite mark evidence using Adobe. Santa Barbara, CA: Forensic Imaging Services, 2000.
15. Dorion RBJ, editor. Bitemark evidence. New York: Marcel Dekker, 2005.
16. Pretty IA, Sweet D. Digital bite mark overlays-an analysis of effectiveness. J Forensic Sci 2001;46(6):385–91.
17. Martin-de las Heras S, Valenzuela A, Ogayar C, Valverde J, Torres JC. Computer-based production of comparison overlays from 3D-scanned dental casts for bite mark analysis. J Forensic Sci 2005;50(1):127–33.
18. Whittaker DK. Some laboratory studies on the accuracy of bite mark comparison. Int Dent J 1975;25(3):166–71.
19. Zhang Z, Monteiro-Riviere NA. Comparison of integrins in human skin, pigskin, and perfused skin: an *in vitro* skin toxicology model. J Appl Toxicol 1997;17(4):247–53.
20. Arheart KL, Pretty IA. Results of the 4th ABFO bitemark workshop-1999. Forensic Sci Int 2001;124:104–11.
21. Swets JA. Measuring the accuracy of diagnostic systems. Science 1998;240:1285–93.
22. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33(1):159–74.

Additional information and reprint requests:
Stella Martin-de las Heras, M.D., Ph.D., B.D.S.
Department of Forensic Medicine and Forensic Odontology
School of Medicine
University of Granada
Avenida de Madrid, 11
18071 Granada
Spain
E-mail: stella@ugr.es