JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, January 2011, Vol. 56, No. S1
doi: 10.1111/j.1556-4029.2010.01602.x
Available online at: onlinelibrary.wiley.com

TECHNICAL NOTE

QUESTIONED DOCUMENTS

Raymond Marquis,[1] *Ph.D.; Silvia Bozza,*[2] *Ph.D.; Matthieu Schmittbuhl,*[3] *Ph.D.; and Franco Taroni,*[1] *Ph.D.*

Handwriting Evidence Evaluation Based on the Shape of Characters: Application of Multivariate Likelihood Ratios*,†

ABSTRACT: A novel Bayesian methodology has been developed to quantitatively assess handwriting evidence by means of a likelihood ratio (LR) designed for multivariate data. This methodology is presented and its applicability is shown through a simulated case of a threatening anonymous text where a suspect is apprehended. The shape of handwritten characters *a*, *d*, *o,* and *q* of the threatening text was compared with characters of the true writer, and then with two other writers, one with similar and one with dissimilar characters shape compared to the true writer. In each of these three situations, 100 draws of characters were made and the resulting distributions of LR were established to consider the natural handwriting variation. LR values supported the correct hypothesis in every case. This original Bayesian methodology provides a coherent and rigorous tool for the assessment of handwriting evidence, contributing undoubtedly to integrate the field of handwriting examination into science.

KEYWORDS: forensic science, handwriting, evidence evaluation, multivariate likelihood ratio, shape analysis, Fourier descriptors

As long as handwriting examiners carry out expertises to give opinions about writership of questioned handwritings, their conclusions are essentially based on both their training and experience (1). The evaluation of findings—sets of similarities and differences observed on questioned and reference handwriting samples—to assess whether the samples are written by the same writer or not is thus subjective. Some studies have already contributed to help examiners to evaluate handwriting evidence, providing occurrences frequency or variability estimation of some handwriting features in given populations (2–8). More recent studies, taking advantage of advances in computing science, performed identification or verification tasks through the quantification of some handwriting features (9–13). An important contribution for forensic scientists was the recent developments made by Srihari et al. (14), who proposed an assessment model designed for writer verification tasks by computing likelihood ratios (LRs), which provide a balanced measure of the degree to which particular items of evidence are capable of discriminating among competing propositions of interest. The model of Srihari et al. (14) takes into consideration a set of different variables, but it assumes that these variables are statistically independent, which is questionable. A new Bayesian methodology has been developed, which is able to handle the complexity of multivariate data (16) that can involve correlated variables. The evidence assessment was performed through the derivation of an LR, by using a model that was an extension of an existing model proposed by Aitken and Lucy (17) in the context of elemental composition of glass fragments. This existing model, designed for multivariate data, assumes two sources of variation (the first between replicates within the same group of fragments, and the second between groups) and a constant variation within groups. However, previous results (15) indicated that a constant variation within writers cannot be assumed: the extent of variability within a writer was shown to be peculiar for each writer. The existing multilevel model was thus developed to accounts for this additional source of variation. The statistical methodology of these developments is fully described in Bozza et al. (16).

The aim of the present paper consists of, after a brief presentation of this Bayesian methodology, demonstrating the application of the model to practical cases through a simulated case of a threatening anonymous text where a suspect is apprehended. Measurements on the threatening text are first compared with reference measurements of the writer of the questioned text and successively with reference measurements of writers who did not write the questioned text. LRs will be derived from these measurements. The examination of the LRs will allow us to determine whether the model correctly supports the hypothesis under which the questioned and the reference measurements were taken from the same writer, or respectively from different writers.

Material and Data

A given individual was asked to copy the printed content of a threatening text (to be considered as the questioned or recovered

[1]Institut de Police Scientifique, School of Criminal Justice, Bâtiment Batochime, University of Lausanne, CH-1015 Lausanne-Dorigny, Switzerland.

[2]Department of Statistics, University Ca' Foscari of Venice, San Giobbe, Cannaregio 873, I-30121 Venezia, Italy.

[3]UMR INSERM 977, Dental Faculty, University Louis Pasteur of Strasbourg, F-67085 Strasbourg, France.

*Presented in part at the 14th Conference of the International Graphonomics Society, September 13-16, 2009, in Dijon, France.

†Supported by Research Grant n. K-11K1-116620 from the Swiss National Science Foundation.

Received 5 May 2009; and in revised form 23 Nov. 2009; accepted 9 Jan. 2010.

material) and to write other given texts in order to obtain contemporaneous reference handwriting material for comparative purposes. Instructions—which were not different to those given to writers in the basic study of this research (15)—had to be followed by the author of the threatening text: handwriting must be usual, without any disguise attempt, on a white page of A4 format with a blue BIC Cristal® ball point pen. The threatening text was written in French—which is the mother tongue of the writer—and contained 15 characters of each letter *a*, *d*, *o,* and *q*, in a text composed of 55 words.

The contours of all closed loops of letters *a*, *d*, *o,* and *q* of the questioned threatening text and the reference handwriting samples were extracted and submitted to the shape analysis methodology described in detail in Marquis et al. (18). Each contour was thus described through a set of variables representing the surface and a series of four harmonics. Each harmonic corresponds to a specific contribution to the shape and is defined by two parameters, amplitude and phase, called Fourier descriptors. The amplitude of a harmonic represents the relative importance of its contribution to the original shape of the contour; the phase represents the orientation of the harmonic contribution. So, the shape of each character can be described through $p = 9$ variables representing, respectively, the surface (S), the amplitude (A_j), and the phase (P_j) of the first four harmonics ($j = 1,...,4$).

The data collected in Marquis et al. (15) were used as the background population to estimate parameters of the statistical model that will be adopted. The surface and the first four pairs of Fourier descriptors were extracted from a large sample of naturally handwritten letters *a*, *d*, *o,* and *q* provided by $m = 13$ writers, with n_i measurements on each writer $i = 1,\ldots,m$.

Statistical Model

The assessment of the value of evidence is performed through the derivation of an LR for multivariate data. The LR measures how the evidence in the particular case alters the odds in favor of a particular proposition. Let us consider the following two propositions of interest:

- H_1: the suspect is the author of the manuscript;
- H_2: the suspect is not the author of the manuscript; an unknown person wrote the manuscript.

A number n_1 of measurements, for the characters of a given letter (*a*, *d*, *o*, or *q*), are performed on the threatening text. These measurements are referred to as the recovered data. A number n_2 of measurements are obtained from manuscripts that were written by the suspect. These measurements are referred to as control data.

A two-level model is implemented, taking into account the within-writer variation and the between-writers variation. Let us denote the recovered and the control replicate measurements by vectors y_1 and y_2. Data are assumed having a multivariate Normal distribution:

$$y_i \sim N(\theta_i, W_i) \tag{1}$$

where $\theta_{1(2)}$ and $W_{1(2)}$ denote the mean vector and the matrix of within-source variances and covariances for the recovered (control) measurements, respectively. For the between-source variation, a Normal distribution is taken for θ, $\theta \sim N(\mu,B)$, where μ denotes the mean vector between sources and B denotes the matrix of between-source variances and covariances. An inverted Wishart distribution is introduced to model the within-source variation, $W_i \sim IW(U,n_w)$.

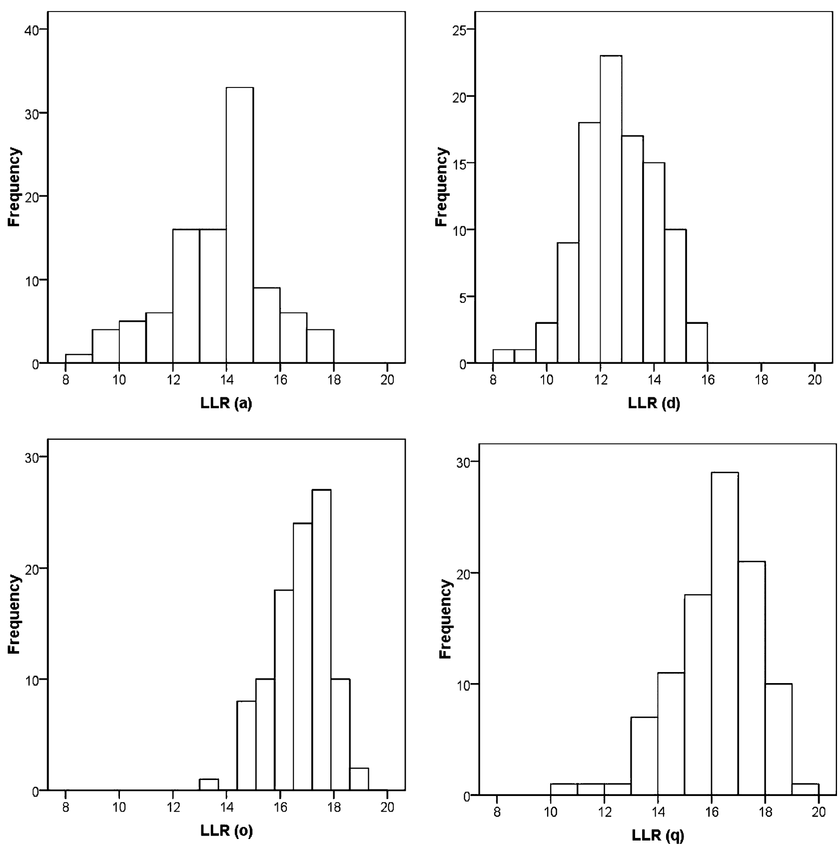

FIG. 1—*Distributions of log-likelihood ratios (LLRs) for each letter a, d, o, and q, where the recovered material and the reference material are written by the same writer.*

Parameters μ, B, and U of the prior densities are estimated from the background data.

The two-level model, which assumes a nonconstant within-source variation, W_i, is in agreement with results presented in Marquis et al. (15) showing that each writer is characterized by a peculiar variation. The model, described in details in Bozza et al. (16), is applied hereafter. For short, the value of the evidence, y_1 (data from the manuscript) and y_2 (data from suspect's material), is the ratio of two probability densities of the form $f\ (y_1,\ y_2\ |H_i)$: one for the numerator, where H_1 is assumed to be true, and one for the denominator, where H_2 is assumed to be true.

Let $\pi\ (\theta,\ W)$ denote the prior density for the unknown parameters θ and W. The value of the evidence is the ratio of likelihoods under the two competing hypotheses:

$$\begin{aligned}\mathrm{LR} &= \frac{f(y_1, y_2|H_1)}{f(y_1, y_2|H_2)}\\ &= \frac{\int f(y_1|\theta, W) f(y_2|\theta, W)\pi(\theta, W) d(\theta, W)}{\int f(y_1|\theta, W)\pi(\theta, W) d(\theta, W) \int f(y_2|\theta, W)\pi(\theta, W) d(\theta, W)}\end{aligned} \tag{2}$$

In the numerator, the mean vectors within writers θ_1 and θ_2 and the matrices of within-writers variances and covariances W_1 and W_2 are assumed to be equal (say to θ and W) but unknown. In the denominator, it is assumed that the source means θ_1 and θ_2, and the matrices W_1 and W_2 are not equal; y_1 and y_2 are taken to be independent as the data are assumed to come from different sources.

If the LR is greater than 1, the evidence supports the hypothesis H_1. If it is lower than 1, then the alternative hypothesis H_2 is supported. Hereafter, results are presented by plotting the logarithm of the likelihood ratio values (LLR) obtained for several samples of measurements. Therefore, positive values support the hypothesis H_1, while negative ones support the hypothesis H_2. The more the LLR differs from 0 (in either direction), the stronger the evidence (19,20).

Results and Discussion

At first, the recovered material was compared to the contemporaneous reference material of the writer who wrote the threatening text. It was attempted to determine whether the measurements made on the recovered and the reference material were able to indicate that the samples were written by the same individual, rather than by two different individuals. A fixed number of 14 characters were randomly drawn from the threatening text and from the reference material of the suspect, respectively, to compute the LR according to the aforementioned methodology. It appeared that the LR value was dependent on the draw; this dependence is related to the natural variation characterizing the handwriting of any given person (the within-writer variability). Therefore, 100 random draws were made for each letter *a*, *d*, *o*, and *q* separately and the LR was computed for each draw. Results, on a logarithm scale, are shown in Fig. 1. Expected positive values were obtained for all the random draws, whatever the letter. Thus, results correctly supported the hypothesis that the same writer wrote the threatening text and the reference documents.

In a second step, the recovered material was compared to reference material coming from a writer who did not write the questioned threatening text. The aim was to determine whether the measurements made on the recovered and the reference materials were able to indicate whether the samples were written by two different writers, rather than by the same one. These experiments were carried out with two different writers (neither of them did in fact write the threatening text), extracted from the available database, say writer 1 and writer 2. At first, the measurements on the threatening text were compared with measurements taken on documents of writer 1. The reason behind the selection of this writer is the high similitude of the shape of his loops with those of the writer of the threatening text (Fig. 2). This similitude was shown in Marquis et al. (15): first, these two writers belonged to the same group of writers according to the general shape of loops. Second, the measurements on the threatening text were compared with measurements taken on documents of the writer 2, chosen for the strong difference in the general shape of his loops with the writer of the threatening text (Fig. 2) according to the findings of Marquis et al. (15). Random draws of characters were made in the same manner as described earlier, but this time the reference material comes from characters of either writer 1 or writer 2. The resulting distributions of LLR values are shown in Figs 3 and 4, respectively. Negative values of LLR were obtained

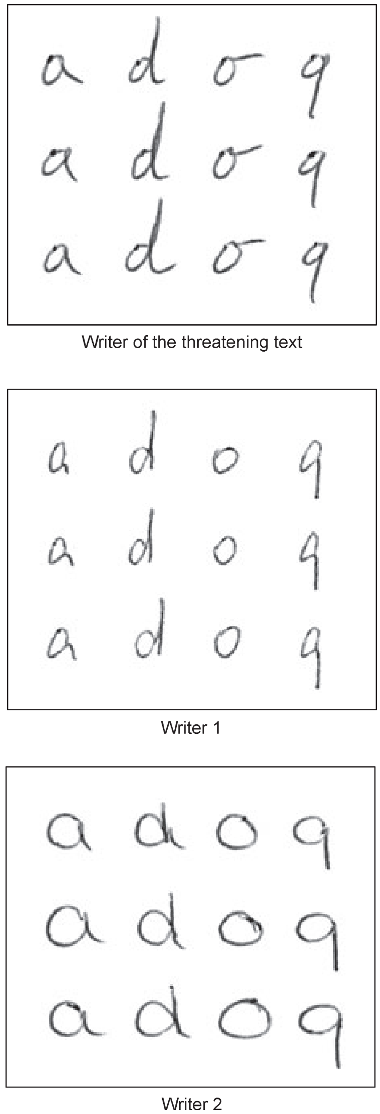

FIG. 2—*Illustration of characters a, d, o, and q of the writer of the threatening text and of the writers 1 and 2, who respectively present a similar and a different general shape of loops compared to the writer of the threatening text.*

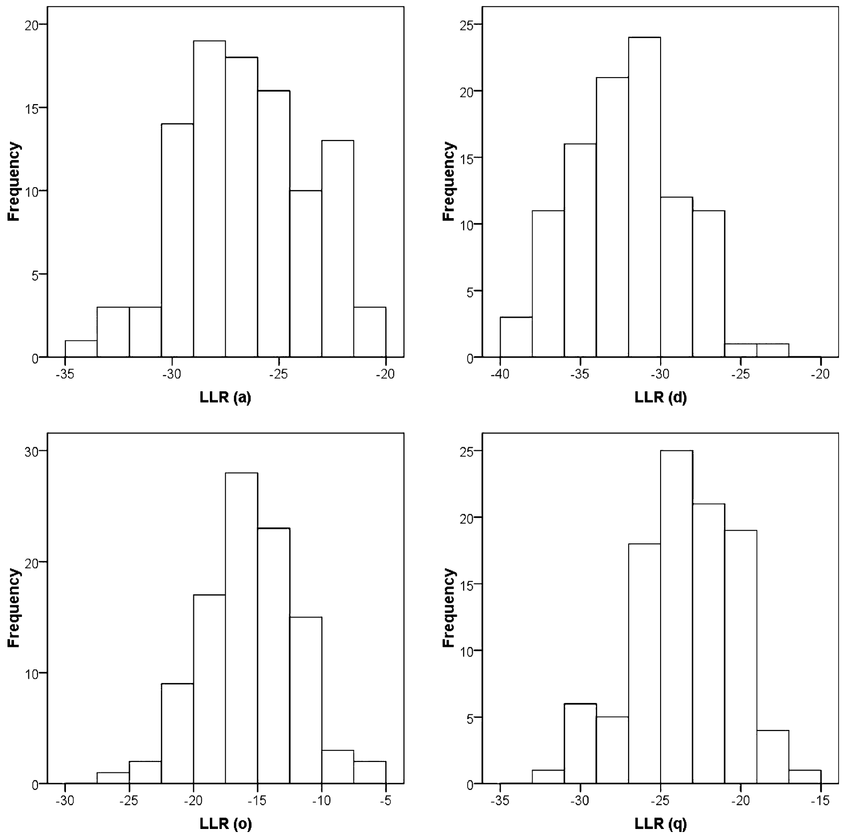

FIG. 3—*Distributions of log-likelihood ratios (LLRs) for each letter a, d, o, and q, where the recovered material is written by the suspect and the reference material by the writer 1 (the first nonwriter), both writers showing a similar general shape of loops according to Marquis et al. (15).*

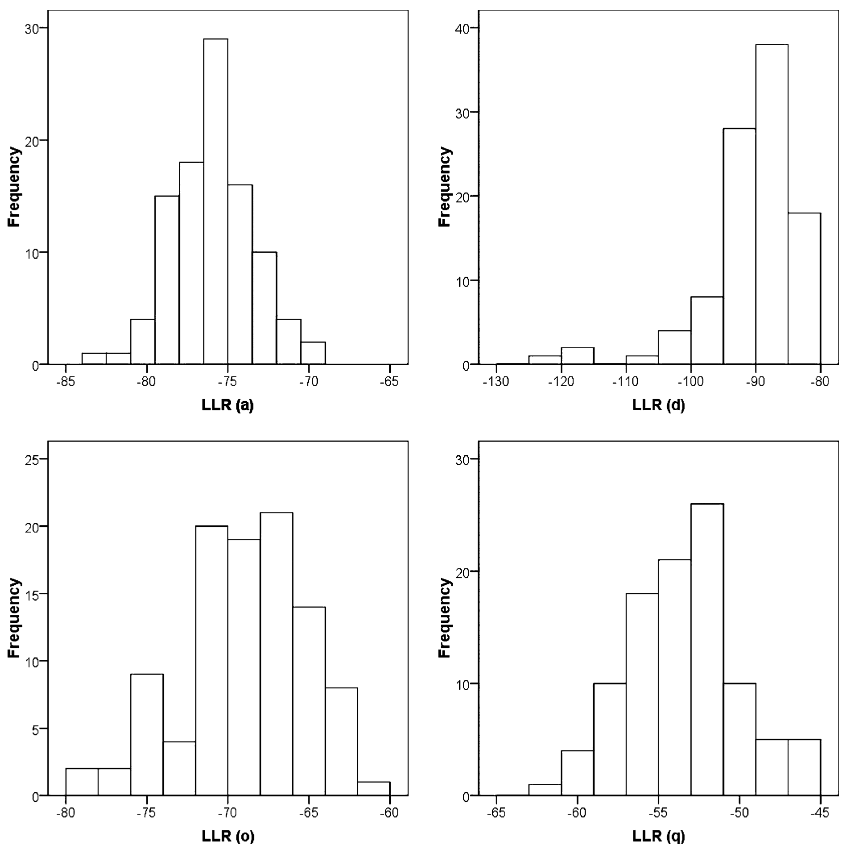

FIG. 4—*Distributions of log-likelihood ratios (LLRs) for each letter a, d, o, and q, where the recovered material is written by the suspect and the reference material by the writer 2 (the second nonwriter), both writers showing a different general shape of loops according to Marquis et al. (15).*

for all results involving either writer 1 or writer 2. Such results were expected as neither writer 1 nor writer 2 wrote the threatening text. The measurements supported the correct hypothesis of nonwritership. However, a marked difference was observed between the results involving the two writers. Indeed, LLR values related to writer 2 were significantly smaller (i.e., more negative) than those associated with writer 1. The hypothesis of nonwritership was thus more strongly supported when the questioned text was compared to the handwriting of writer 2 than to the handwriting of writer 1. As the shape of letters of writer 2 was more different from that of the questioned text compared to the shape of letters of writer 1, these findings show that the more different the reference material from the questioned text, the stronger the value of evidence support the hypothesis that the author of the reference material did not write the questioned text. From another point of view, if the writer shows a shape of loops more similar to that found in the threatening text, then it is more likely to get false-positive values. In particular, LR greater than 1 has higher probability of occurring with writer 1 than with writer 2. However, in the case at hand, no positive value—which would have corresponded to false-positive results—was obtained whatever the writer and the letter *a*, *d*, *o,* and *q* involved.

These findings show that the proposed methodology represents a valid tool to properly handle the shape variation and to support the correct hypothesis either when H_1 is true or when H_2 is true. It must be emphasized that in situations where H_2 is true, the model even works if the general aspect of shape is very similar between the questioned document and the reference material.

Note that the LLR results are only indicative. The background data—necessary to choose the prior densities for the unknown parameters (16)—are made of measurements taken on handwriting samples collected from only 13 individuals. This has an influence on the magnitude of the LR value. Furthermore, the model necessitates some adjustments to make it possible to handle different letters simultaneously, taking into consideration the correlation between letters which can be different from writer to writer (15,16). Nevertheless, the statistical model developed and presented here was successfully applied to multivariate data, such as those extracted from the shape of characters *a*, *d*, *o,* and *q*, in order to assess whether the considered measurements support the hypothesis of writership or the alternative hypothesis. This novel Bayesian methodology provides a coherent and rigorous tool—represented by a quantitative support given by the LR value—to help handwriting examiners to assess handwriting evidence. This must undoubtedly be considered as a notable major step to integrate the field of handwriting examination into science. Finally, the auspicious findings of this research suggest the applicability of this approach to other fields of forensic science, where multivariate data are becoming more and more prevalent.

Conflict of interest: The authors have no relevant conflicts of interest to declare.

References

1. Lee F. Review of "handwriting identification evidence in the post-Daubert world." J Am Soc Questioned Doc Examiners 1998;1:67–8.
2. Crane A. The frequency of round handwriting in Edmonton, Alberta Schools. Société Canadienne des Sciences Judiciaires 1999;32:169–74.
3. Horton R. A study of the occurrence of certain handwriting characteristics in a random population. Int J Forensic Doc Examiners 1996;2:95–102.
4. Livingston O. Frequency of certain characteristics in handwriting pen-printing of two hundred people. J Forensic Sci 1963;8:250–9.
5. Muehlberger R, Newman K, Regent J, Wichmann J. A statistical examination of selected handwriting characteristics. J Forensic Sci 1977;22:206–15.
6. Wing A, Nimmo-Smith I. The variability of cursive handwriting measure defined along a continuum: letter specificity. J Forensic Sci Soc 1987;37:297–306.
7. Zimmerman J. Frequency of omission, perseveration and reversals in handwriting and relative value as individual characteristics. J Forensic Ident 1990;40:127–34.
8. Zimmerman J. Counting handwriting characteristics. Int J Forensic Doc Examiners 1998;4:318–22.
9. Bensefia A, Paquet T, Heutte L. A writer identification and verification system. Pattern Recognit Lett 2005;26:2080–92.
10. Leedham G, Chachra S. Writer identification using innovative binarised features of handwritten numerals. Proceedings of the 7th International Conference on Document Analysis and Recognition; 2003 Aug 3-6; Edinburgh, Scotland. International Association of Pattern Recognition, 2003;413–6.
11. Plamondon R, Lorette G. Automatic signature verification and writer identification—the state of the art. Pattern Recogn 1989;22:107–31.
12. Said H, Tan T, Baker K. Personal identification based on handwriting. Pattern Recogn 2000;33:149–60.
13. Schomaker L, Franke K, Bulacu M. Using codebooks of fragmented connected-component contours in forensic and historic writer identification. Pattern Recognit Lett 2007;28:719–27.
14. Srihari S, Huang C, Srinivasan H. On the discriminability of the handwriting of twins. J Forensic 2008;53:430–46.
15. Marquis R, Taroni F, Bozza S, Schmittbuhl M. Quantitative characterization of morphological polymorphism of handwritten characters loops. Forensic Sci Int 2006;164:211–20.
16. Bozza S, Taroni F, Marquis R, Schmittbuhl M. Probabilistic evaluation of handwriting evidence: likelihood ratio for authorship. Appl Stat 2008;57:329–41.
17. Aitken CGG, Lucy D. Evaluation of trace evidence in the form of multivariate data. Appl Stat 2004;53:109–22.
18. Marquis R, Schmittbuhl M, Mazzella WD, Taroni F. Quantification of the shape of handwritten characters: a step to objective discrimination between writers based on the study of the capital character O. Forensic Sci Int 2005;150:23–32.
19. Good IJ. Probability and the weighing of evidence. London, UK: Griffin, 1950.
20. Robertson B, Vignaux GA. Interpreting evidence: evaluating forensic science in the courtroom. Chichester, UK: John Wiley & Sons, 1995.

Additional information and reprint requests:
Raymond Marquis, Ph.D.
Institut de Police Scientifique
School of Criminal Justice
Bâtiment Batochime
University of Lausanne
CH-1015 Lausanne-Dorigny
Switzerland
E-mail: raymond.marquis@unil.ch