Limited Populations – Are They Feasible for Handwriting Examinations?

Chris Anderson, BSc[1] and Julian Leslie, PhD[2]

[1] Chris Anderson & Co Pty Ltd, Forensic Document Examiners, P.O. Box 2778 Carlingford, NSW 2118, Australia
[2] Associate Professor, Statistics, DEFS, Macquarie University, Sydney, NSW 2109, Australia

The term *limited population* refers to an examination involving a small number of potential writers who could have written a questioned document containing a very limited number of characters. An investigation was undertaken to ascertain the impact a limited population of writers would have on identifying the actual writer. Two simple statistical models were used in this study. Unsurprisingly, the chance of identifying the writer reduces as the number of potential writers increases. However, this reduction became quite marked for certain values of quantities used in the models. The quantities of relevance were the number of key distinguishing features in the writing and 2 probabilities—1 associated with incorrectly deciding that a character was not written by the writer of the questioned document and the other associated with correctly identifying a character as not written by a person who did not write the questioned document.

Introduction

The term *limited population* refers to an examination where it was claimed that the number of potential writers who could have written a questioned document was small and of known size. In addition, there were only a few characters of writing in the questioned document (including letters, numerals, symbols, punctuation, spacing, positioning, and height ratios). The use of limited populations in forensic handwriting examinations has evoked intense discussion among forensic document examiners (FDEs) at professional meetings where 2 schools of thought were evident. Opponents claim that examinations involving limited populations should not be contemplated because of the uncertainty surrounding the number of potential writers as well as the small number of characteristics on which to base an identification. Proponents state that there can be special circumstances when examinations involving limited populations are appropriate.

Looking at the problem from a mathematical standpoint assists in determining which issues need to be thought about and if such an examination should be considered. By formulating the problem in terms of the probabilities of certain events occurring, a formal structure can be imposed on the problem. The advantage of this would be that it could be shown how the elements in a model influence the outcome (i.e., the chance that a particular writer wrote the document given the match information). As might be expected, a key element was the number of potential writers of the questioned writing. While recognizing that more sophisticated models could be proposed, the models used in this study produce interesting effects. This research is not to be viewed as support for or rejection of the use of limited populations but rather to clarify the influence of the number of writers.

In general, without any claims of a known small number of potential writers and a very limited amount of questioned writing, the FDE often may form an inconclusive opinion or a very weak conclusion regarding the authorship of questioned writing.

The suggestion of narrowing the population base would normally come from the client who instructs the FDE to assume that there were only a specified number of potential writers. This approach may be treated by some FDEs with a certain degree of scepticism, as it can be perceived as a device to artificially increase the chance of implicating a particular person when there seems to be little or no justification for reducing the population base of potential writers. This study was devised to untangle some of the confusion surrounding this issue and quantify the effect of population size on the likelihood that a particular individual wrote a questioned writing.

Methods and Materials

This study was based on an actual case where a small piece of questioned text contained only 10 characters. It was reported that 1 of 5 suspected/potential writers wrote the text. Known writings of the 5 individuals were provided that included normal course-of-business and request exemplars. During the examination, a number of differences were observed between the questioned and specimen writing for 4 writers (Table 1).

Writer A	Had no differing characters
Writer B	Had 2 differing characters
Writer C	Had 2 differing characters
Writer D	Had 2 differing characters
Writer E	Had 4 differing characters

Table 1. The number of differences observed between the questioned and known writing for the potential writers of the questioned writing.

To assist with determining the chance that a particular writer wrote the questioned writing, 2 statistical models were constructed. **Model 1** calculated the probability of *Writer A* having written the questioned entries based on the information that 4 writers, other than *Writer A*, had at least 1 observable difference between their writing and the questioned writing. **Model 2** incorporated all the information of the observed differences in Table 1 to calculate the probability that *Writer A* wrote the questioned writing. The point of considering **Model 1** was that it included all the outcomes mentioned in the case study but allowed the discussion to be extended to an arbitrary number of writers, whereas **Model 2** was specific to the actual number of writers used in the study. To extend the study using **Model 2** and include further writers, it was necessary to take a random sample of a number of feasible differences that an FDE might encounter with populations of 10 and 15 writers. In other words, a number of simulations were produced that contained typical differences which 10 or 15 writers might generate. In applying the models, these features were present:

1. The questioned writing contained 10 characters;
2. There were only 5 possible writers of the questioned writing;
3. One of those 5 writers wrote the questioned writing;
4. The writing of the 1st writer (*Writer A*) had no perceived differing characters to the questioned writing, and the writing of the remaining 4 writers had at least 1 differing character (Table 1); and
5. All writers provided comparable amounts of specimen writing being reasonably extensive and consisting of their normal course-of-business writing.

It was noted that in a small amount of questioned writing an *apparent difference* might not be assessed as a *difference* had there been a large base of writing for comparison. Therefore, any discussion of differences includes *perceived differences*.

The formation of both models was centered upon 2 probability parameters, $\boldsymbol{p}$ and $\boldsymbol{\pi}$. Parameter $\boldsymbol{p}$ was defined thus: "Suppose *X* was the person who *wrote* the questioned writing, then define $\boldsymbol{p}$ to be the chance that *X* wrote a character that the FDE (erroneously) concluded, based on *X's* specimen writing, that 'this character was not written by *X*' (misidentification)." Similarly, parameter $\boldsymbol{\pi}$ was defined: "Suppose *Y* was the person who *did not write* the questioned writing; i.e., *Y* was not *X*. Then $\boldsymbol{\pi}$ is the chance that the FDE (correctly) concluded that a character, in fact written by *X*, was not written by *Y*, based on *Y's* specimen writing."

A technical note that may be of interest to statisticians is that in determining $\boldsymbol{p}$, it was the random variation in *X's* writing of a character that was considered in assessing whether that character was written by *X*; whereas, in the case of $\boldsymbol{\pi}$, it was the randomness of *Y's range of variation* of that character and whether that range failed to include the questioned character (this being the criteria used in assessing whether *Y* did not write the character).

The expectation was that $\boldsymbol{p}$ would be much smaller than $\boldsymbol{\pi}$ because 1st, in the case of $\boldsymbol{p}$, the event of interest was that *X* had written a character sufficiently outside their normal range of variation for that character to be identified as significantly different. Second, in the case of $\boldsymbol{\pi}$, the event of interest was that a random individual had a range of variation for a character that failed to include the observed character written by *X*. In other words, the *X*-written character was a

given, and π was the chance that a random individual would have a range of variation for that character that failed to include the *X*-written character. This assessment was dependent upon there being sufficient normal-course-of-business writing for *X* so it would be seen to be unusual for *X* to write outside that range; i.e., p was small. In plain terms, p was concerned with *within-person variation,* whereas π was concerned with *person-to-person variation. Person-to-person variation* in the writing of a character would be expected to be much greater than the *within-person variation*, leading to π being large compared to p.

Both models assumed that each character had the same p, and further, that each character and each (non-X) writer had the same π; i.e., all characters had the same value of π for each of the nonwriters of the questioned writing. It might be argued that some characters were more complex and therefore easier to detect as differing. Hence these characters should have different probabilities (weightings) associated with them. In addition, different FDEs' assessment techniques and/or experience could lead to differing probabilities as well. More sophisticated models would attempt to take these factors into consideration; however, this was beyond the scope of this present work.

Regarding common authorship, Hilton (1982) stated, "If two writings are by a single person, then no fundamental differences should exist." If this were factual, then p should be 0. However, that remark was predicated on there being a sufficient quantity of both questioned and specimen writing to be able to establish that a feature of the writing was in fact fundamentally different. The use of p to represent uncertainty, in effect, allowed for a margin of error in the FDE's assessment of each character in the questioned writing.

In a *limited population* there may only be 1 formation of a particular letter or character in a questioned writing that would make it very difficult to determine whether any observed inconsistency between the questioned and specimen writing was a fundamental difference or just a variant or accidental (Harrison 1966). In reality, the FDE may never be able to assess whether such an inconsistency was a fundamental difference or not. For this exercise, any inconsistency that would lead the FDE to suspect that there may be a difference was deemed to be a significant

Let n = the number of writers

Let x = the number of characters in the questioned writing

Let p = the probability of detecting a significant difference by writer of the questioned writing

Let π = the probability of detecting a significant difference by writer who did not write the questioned writing

$$\Pr(A \text{ wrote} \mid A = 0 \text{ and } B, C, D, \ldots \geq 1) = \frac{\Pr(V)}{\left[\Pr(V) + \left(\Pr(W) \times (n-1)\right)\right]} \quad (1)$$

where

$$\Pr(V) = \Pr(A = 0 \text{ and } B, C, D, \ldots \geq 1 \mid A \text{ wrote}) = (1-p)^x \left[1 - (1-\pi)^x\right]^{n-1} \quad (2)$$

and

$$\Pr(W) = \Pr(A = 0 \text{ and } B, C, D, \ldots \geq 1 \mid B \text{ wrote}) = (1-\pi)^x \left[1 - (1-p)^x\right]\left[1 - (1-\pi)^x\right]^{n-2} \quad (3)$$

Equations 1, 2, and 3.

$$\Pr(B\ wrote \mid A=0; B, C, D, \ldots \geq 1) = \frac{\Pr(W)}{\left[\Pr(V)+\left(\Pr(W)\times(n-1)\right)\right]}$$

Equation 4.

$$P_1 = \Pr(A=0\ and\ B=2 \mid A\ wrote) = (1-p)^{10} \times {}^{10}C_2 \times \pi^2 \times (1-\pi)^8$$

Equation 5.

$$P_2 = \Pr(A=0\ and\ B=2 \mid B\ wrote) = (1-\pi)^{10} \times {}^{10}C_2 \times p^2 \times (1-p)^8$$

Equation 6.

$$\Pr(X) = \Pr(A \text{ wrote} \mid\ A=0\ \&\ B=2) = \frac{P_1/2}{(P_1+P_2)/2} = \frac{2.3593\times10^{-3}}{2.5068\times10^{-3}} = 0.9412$$

Equation 7.

$$\Pr(Y) = \Pr(B \text{ wrote} \mid\ A=0\ \&\ B=2) = \frac{P_2/2}{(P_1+P_2)/2} = \frac{1.4746\times10^{-5}}{2.5068\times10^{-3}} = 0.0588$$

Equation 8.

difference—the implication being that the chance of the FDE identifying a genuine difference for a character when the person was not the actual writer (i.e., π) can still be rather small. This also raises the problem of differential amounts of specimen (or known) writing of the various potential writers. Clearly, if the prime suspect had voluminous amounts of specimen writing available, it might be possible to show that every character had occurred somewhere, falling within the known range of variation for that writer. If, however, another writer provided only a limited amount of specimen writing, *perceived differences* may arise simply because the range of variation for that writer had not been adequately determined.

Results

For **Model 1**, the conditional probability that *Writer A* wrote the questioned writing (given the specimen writing of *Writer A* had no significant differences and the specimen writing of the 4 other writers had at least 1 significant difference) was expressed in Equations 1, 2, and 3.

In Equation 1, the expression Pr*(A wrote | A = 0 and B, C, D,...≥1)* was to be interpreted as the *conditional probability* of *Writer A* having written the questioned writing given the observations. The observations consisted of the events that *Writer A* had no differences, while writers *B, C,* etc. had at least 1 character difference. For statisticians, the term *conditional probability* has a specific notation and meaning. The notation Pr*(A | B)* means "the probability that event A occurs given that event B has occurred," and the definition is Pr*(both events A and B occur)*/Pr*(event B occurs)* where Pr means "probability." Further, Pr*(V)* means the *conditional probability* of the observations given that *Writer A* wrote the questioned writing, and Pr*(W)* means the *conditional probability* of the observations given that *Writer B* wrote the questioned writing, respectively.

The formula used in Equation 1 is known as Bayes formula or rule (Moore and McCabe 2003), where an *a priori* probability (i.e., the probability that any particular writer was the true writer of the questioned writing) was assumed to be $1/n$ where n = the number of writers. To put it another way, it was assumed, in the absence of any other information, that before looking at the questioned document each of the n writers was equally likely to have been the writer of the questioned document.

To give some sense of what these formulae deliver, take the values $p = 0.2$; $\pi = 0.5$; $n = 5$; $x = 10$ into Equations 1, 2, and 3. Then the *conditional probability* of *Writer A* having written the questioned writing given the observations was calculated to be Pr*(A wrote | A = 0 and B, C, D,...≥1) = 0.9685*—meaning there was approximately a 97% chance that *Writer A* wrote the questioned writing given that all writers, apart from *Writer A,* had at least 1 observable difference.

Conversely, using the same values as above, the probability of *Writer B* having written the questioned writing can be calculated as in Equation 4. The *conditional probability* of *Writer B* having written the questioned writing, given the observations, was considerably smaller and calculated to be Pr*(B wrote | A = 0 and B, C, D,...≥1) = 0.007863.* The same probability values apply for writers *C, D,* and *E* having written the questioned writing. Also calculated was the analogous probability in the case of populations of 10 and 15 writers. A list was made (Table 2) that calculated the probability of *Writer A* having written the questioned writing based on having a population of 2, 5, 10, and 15 writers. In addition, a plot (Figure 2) was prepared of these calculated probabilities. Not surprisingly, as the number of writers increased so the probability of *Writer A* having written the questioned writing decreased.

Model 1	
	probability
2 writers	0.9919
5 writers	0.9685
10 writers	0.9319
15 writers	0.8979

Table 2. Model 1 represents the *conditional probabilities* of *Writer A* having written the questioned writing based on populations of 2, 5, 10, and 15 writers.

To understand the mathematical formulation of **Model 2**, the simple case of just the first 2 writers, A and B, was considered. In this case Writer A had no characters differing significantly from the specimen writing, whereas *Writer B* had 2 characters which differed significantly. In the absence of any other information, the chance that

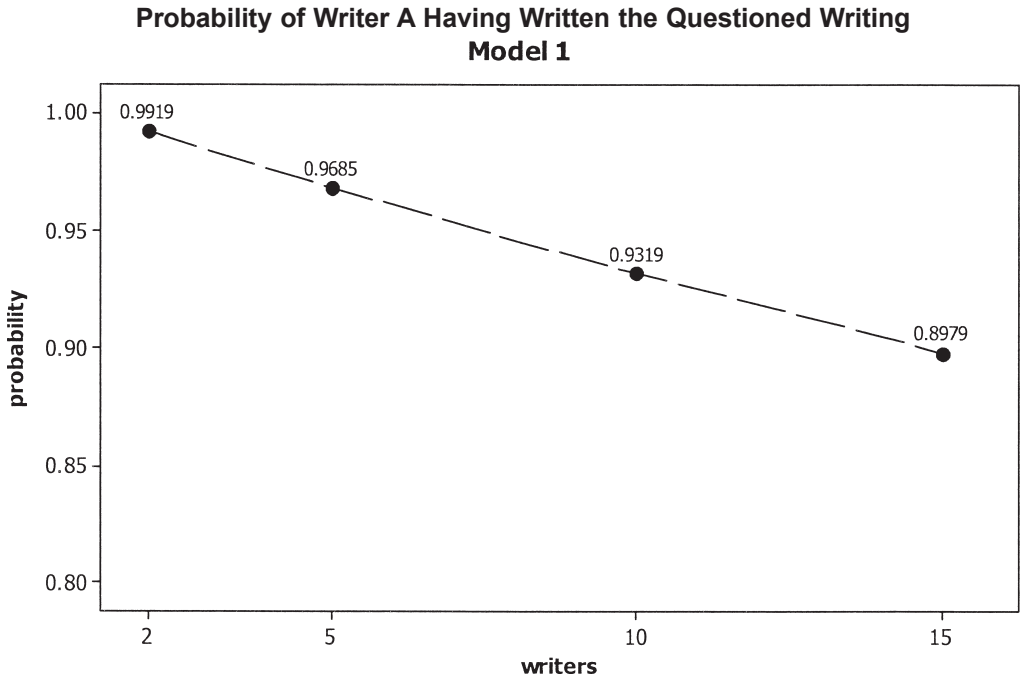

Figure 1 . Model 1 plot of *conditional probabilities* of *Writer A* having written the questioned writing based on populations of 2, 5, 10, and 15 writer.

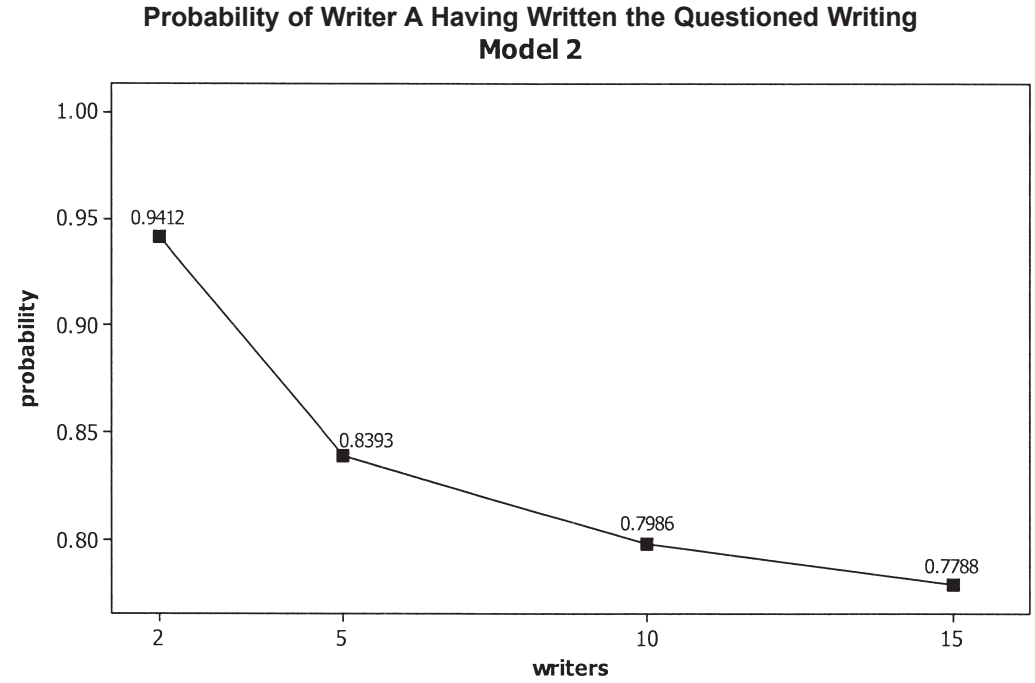

Figure 2. Plot of **Model 2** *conditional probabilities* of *Writer A* having written the questioned writing based on populations of 2, 5, 10, and 15 writers.

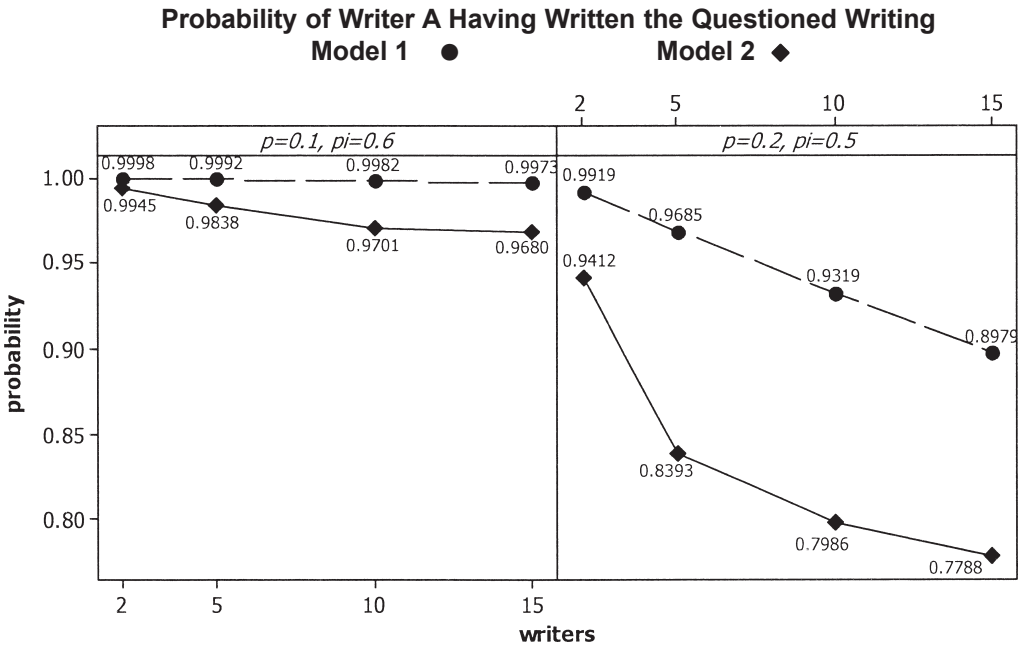

Figure 3 . Panel plot of **Models 1** and **2** showing the *conditional probabilities* of *Writer A* having written the questioned writing based on populations of 2, 5, 10, and 15 writers when $\boldsymbol{p} = 0.1$, $\boldsymbol{\pi} = 0.6$ and $\boldsymbol{p} = 0.2$, $\boldsymbol{\pi} = 0.5$.

Writer A wrote the writing was *0.5* in view of the fact that there were just 2 individuals and it was assumed that 1 of them did write the questioned writing. Consider the chance P_1 that *Writer A* had no characters differing and *Writer B* had 2 characters differing given *Writer A* wrote the questioned writing. This chance can be expressed using the mathematical expression in Equation 5.

In Equation 5, ${}^{n}C_{r}$ represents the number of ways to choose r things from n distinct things, in particular ${}^{10}C_{2} = 45$. Similarly, the chance that *Writer A* had no characters differing and *Writer B* had 2 characters differing given that *Writer B* wrote the questioned writing was expressed in a similar manner to that of Equation 5.

The same $\boldsymbol{p}$ and $\boldsymbol{\pi}$ values used in **Model 1** were applied to this model; i.e., $\boldsymbol{p} = 0.2$; $\boldsymbol{\pi} = 0.5$. Thus, by Bayes formula, the chance that *Writer A* wrote the questioned writing given that *Writer A* had no characters differing and *Writer B* had 2 characters differing was expressed in Equation 7.

By including the information on the observed differences**,** the probability that *Writer A* wrote the questioned writing increased from *0.5* to *0.94* in this case of just 2 writers. The probability that *Writer B* wrote the questioned writing given the information of the differences was calculated in a similar manner to those shown in Equations 7 and 8. This probability was considerably less than the probability calculated for *Writer A*.

Using an extension of these combinatorial expressions to incorporate the situation described in Table 1 (where there were writers with varying numbers of similarities and differences), the probability that Writer A wrote the questioned writing given this information can be calculated. As the combinatorial expressions become long and complex, they are omitted. By simulating a random set of differences for the extra writers, the average probability that *Writer A* wrote the questioned writing when there were 10 and 15 writers in the population was also calculated. The simulation was based on a binomial distribution of $\boldsymbol{n} = 10$ (where $\boldsymbol{n}$ = the number of characters) and $\boldsymbol{p} = 0.5$, assuming that the writers who did not write the questioned writing have a chance of 0.5 of having their range of variation for a particular character include that questioned character. The average of all the calculated probabilities for 10 and 15 writers was derived from

10 such simulations. The calculated conditional probabilities that *Writer A* wrote the questioned writing based on populations of 2, 5, 10, and 15 writers was established (Table 3).

Model 2	
	probability
2 writers	0.9412
5 writers	0.8393
10 writers	0.7986
15 writers	0.7788

Table 3. Model 2 representing conditional probabilities of *Writer A* having written the questioned writing based on populations of 2, 5, 10, and 15 writers.

A list (Table 4) of the *conditional probabilities* that *Writer B* wrote the questioned writing for 2, 5, 10, and 15 writers was calculated based on **Model 2**. These probabilities were very small and decreased as the number of writers increased.

Model 2	
	probability
2 writers	0.0588
5 writers	0.0525
10 writers	0.0499
15 writers	0.0487

Table 4. Model 2 representing the *conditional probabilities* of *Writer B* having written the questioned writing based on populations of 2, 5, 10, and 15 writers.

Discussion of Results

Both models calculate probability values that *Writer A* wrote the questioned writing given the information about the similarities and differences. If the necessary criteria are satisfied, then there is no reason for not using the calculated probabilities in casework. With expanding knowledge and information, there will be an increasing ability to refine the values for p and π to more realistic values—with specific values for individual characters and individual subjects. It is important to note that the values used in this study were deliberately chosen to be conservative estimates—*conservative* in the sense of being fair to all writers. A smaller value of p and a larger value of π would lead to higher posterior probabilities (i.e., the probabilities calculated).

The 2 parameters p and π were not probabilities of a FDE determining whether a character was written by *X* or by *Y*. There was no intention to imply that FDEs were no better than 50% accurate in assessing whether a character was deemed to be a difference or just a variant. The point here was that when the information available to the FDE was very limited (both in terms of the number of characters in the questioned writing and in terms of the amount of specimen writing available), then using 50% accuracy may not be unreasonable.

The change in posterior probabilities can be seen when the values for p and π were altered to *0.1* and *0.6,* respectively. This makes it less likely that the FDE will make an error when assessing the true writer and less likely that the FDE will make an error when comparing the questioned characters with the specimen writing of someone who did not write the questioned characters. Both error probabilities (p and $1-\pi$) have decreased—which would be the case if more specimen writing became available for all writers. The probability that *Writer A* wrote the questioned writing under both models changed; that is, the posterior probabilities increased. A list of the probabilities was calculated (Table 5) when $p = 0.1$ and $\pi = 0.6$ for **Models 1** and **2** for populations of 2, 5, 10, and 15 writers. The previously calculated probabilities were included for **Models 1** and **2** when $p = 0.2$ and $\pi = 0.5$ by way of comparison.

A panel plot (Figure 3) of these data shows graphically the increased posterior probabilities in using the lower p value and a higher π value on the probability that *Writer A* wrote the questioned writing for **Models 1** and **2**.

An obvious question arose as to which model was the more appropriate to use. On the 1 hand, it was desirable to use the specific information about the numbers of differences exhibited by each writer when that was available—this typically being the case. However, when it was of interest to see the effect of increasing the numbers of writers in a given case, both methods allowed for some indication of this effect to be

	Model 1	*Model 2*	*Model 1*	*Model 2*
	$p = 0.1$ $\pi = 0.6$		$p = 0.2$ $\pi = 0.5$	
2 writers	0.9998	0.9945	0.9919	0.9412
5 writers	0.9992	0.9838	0.9685	0.8393
10 writers	0.9982	0.9701	0.9319	0.7986
15 writers	0.9973	0.9680	0.8979	0.7788

Table 5. Models 1 and **2** showing *conditional probabilities* of *Writer A* having written the questioned writing based on populations of 2, 5, 10, and 15 writers when $p = 0.1$, $\pi = 0.6$ and $p = 0.2$, $\pi = 0.5$.

obtained. Both models calculated the probability that *Writer A* wrote the questioned writing given the information of the similarities and differences. **Model 1** used 1 writer with no significant differences and all other writers having at least 1 observable difference, which could be advantageous when extrapolations were required on increasing the number of writers. **Model 2** calculated the probability of 1 writer having no differences and the other writers having a defined number of differences.

Conclusions

This study clearly established a very strong case for implicating *Writer A* as the writer of the questioned writing. Both models assumed that the characters were treated equally. In reality, not all characters have equal weight, some having far more importance in identifying or excluding a potential writer than others. There can be difficulty in establishing that a *perceived difference* was a *fundamental difference* when dealing with small numbers of characters in a *limited population* examination—this being further reason to use conservative values for p and π.

Another matter to consider (and 1 that relies on the expertise of the FDE) should be when to decide it is appropriate to involve a *limited population*. Before embarking on such an examination, the FDE should assess whether the writing is identifiable despite the small number of characters for comparison. Here the usual features present in the writing in combination should be assessed, including the speed and fluency of writing, the complexity of the writing, and its perceived individuality.

There must be sufficient representative writing of all potential writers. The importance of this cannot be overemphasised. As previously discussed, all such writings should be of a similar quantity and written over a similar range of time to ensure good representation of each individual's writing habits and to ensure that the writer's range of variation can be properly and adequately assessed. These requirements ensure that the examination is not skewed towards a particular writer. Request writings need to be used with care, as they may not be a representative sample.

The group of writers considered in this exercise included the person who actually wrote the questioned writing. All the calculations were based on this premise. This begs the question, how can it be ensured that this criterion is satisfied in casework? The stark truth is that it cannot be known. Hence, it has to be assumed that one of the writers of the group wrote the questioned writing. However, circumstances of the case can make this assumption perfectly justifiable, but not always. It is extremely important that any assumption or assumptions relied upon by the FDE in arriving at a conclusion must be clearly and openly stated in the report, together with a statement that if the assumption is found at some future time not to be justified, then any conclusions based thereon will need to be reviewed.

FDEs have been intuitively aware that with only a few features of writing on which to base an identification, an increased number of potential

writers may decrease the chances of finding the actual writer of the questioned writing. This effect was clearly apparent with the plots showing decreasing probabilities calculated for 2, 5, 10, and 15 writers when considering the conditional probability that *Writer A* wrote the questioned writing.

In conclusion, in the special circumstance of a *limited population* examination where all the conditions defined for the aforementioned models were satisfied, then it would be possible to proceed with a handwriting examination and ultimately arrive at a worthwhile conclusion about the authorship of questioned writing—provided the posterior probabilities were large enough.

References

Harrison, W. R. (1966). *Suspect Documents Their Scientific Examination*. Second Impression with Supplement. London: Sweet & Maxwell Ltd., pp. 288-372.

Hilton, O. (1982). *Scientific Examination of Questioned Documents*. Revised ed. New York: Elsevier North Holland, Inc., pp. 153-171.

Moore, D. S. and McCabe, G. P. (2003). *Introduction to the Practice of Statistics*. 4th ed. New York: W. H. Freeman and Co.