Research Article

International Journal for Public Policy, Law and Development
Volume 1, Issue 1, Year 2024

Beyond Certainty: Statistical Pitfalls in Forensic Signature Analysis

Alexander Plant[1], Aziz Ben Jemia[2], Stephen Foster[3]

[1]University College of London, USA
[2]University of San Francisco, USA
[3]Penn State York, USA

Corresponding Author:
Aziz Ben Jemia
University of San Fransisco
E-mail: benjemiaa21@gmail.com

Received: 07-12-2024, Manuscript No. JQR/IJPLD/13; **Editor Assigned**: 08-12-2024, Manuscript No. JQR/IJPLD/13; **Reviewed:** 22-12-2024, Manuscript No. JQR/IJPLD/13; **Published:** 31-12-2024

ABSTRACT:

In this commentary, we examine forensic handwriting analysis, focusing on statistical issues that arise in expert probability judgments. We highlight the challenges in accurately assessing forgery probability, especially as factors such as population size and the base rate of counterfeiting skills vary. In urban and global population contexts, even small error rates in signature matching can lead to high replication likelihood, questioning the typical high confidence expressed by forensic experts. Furthermore, we explore how base-rate assumptions shift within populations of professional counterfeiters, who significantly increase the chances of signature replication. These findings argue for more nuanced approaches in expert testimony, emphasizing a need for improved statistical frameworks and clearer communication to ensure judicial accuracy and avoid potential misinterpretation.

KEYWORDS: Forensic, United States, Inconclusive

1. INTRODUCTION

Forensic signature analysis is a commonly accepted tool found in many legal cases in the United States—until the Daubert case in 1995, handwriting expert evidence was seen as nearly infallible and was trusted whole-heartedly by the courts. Despite the longstanding use of forensic handwriting analysis, recent studies on the errors found in such handwriting experts (e.g., Crot & Marquis, 2024; Kichlin et al., 2022; Martire et al., 2018) have raised questions about how much weight, so to speak, should be placed on forensic evidence of this nature. Considering the implications of expert accuracy, it is of critical importance to understand the way forensic signature evidence should be framed and understood by the courts so that its impact can be effectively weighed by jurors and court officials in court hearings.

There are various contributors to potential "mistakes" forensic signature analysts may make in their judgments—these "mistakes" are, perhaps, more accurately characterized as false match rates, indicating that an analyst has claimed two signatures to be written by the same individual when, in reality, one of the signatures is forged by another person. Kang et al (2023) recently investigated the factors which might lead to false match rates, finding that error rates were higher when no peer review of responses was incorporated, and even found that "inconclusive" responses were less likely to occur when a monetary incentive was introduced. In a test of the forensic confirmation bias, Kukucka & Kassin (2014) found that analysts were more likely to claim a match to the perpetrator when they had knowledge of a perpetrator's

International Journal for Public Policy, Law and Development
Volume 1, Issue 1, Year 2024

confession in the case, suggesting that perceptions of guilt can somehow contribute to whether or not analysts make claims about signature matching. Finally, perhaps unsurprisingly, difficulty scores did increase the likelihood of false matches in an empirical study by Lee and colleagues (2006).

A final area of research has tried to use statistical methods to understand where gaps in understanding and application of forensic signature evidence may lie. Broadly speaking, forensic evidence should be subjected to various statistical considerations which are applied to other forms of scientific evidence—as noted by Kafadar (2014). These considerations include validity, consistency, logical error probabilities, positive/negative predictive values, and estimate uncertainty. As demonstrated by Kafadar (2014), these factors can be approximated using simulated scenarios which mirror real-world contexts. This can be particularly helpful as empirical work may not accurately represent the real-world applications of forensic signature evidence in the court system.

The current study sought to investigate potential issues in the application of forensic signature evidence by exploring various features of the forensic evidence paradigm. These potential issues are based on the manner in which forensic signature evidence is presented in most court systems—the expert stating a high level of confidence that a signature is/is not a match to a genuine source. More specifically, we will use probability to estimate 1) the likelihood of at least a single replicator/no genuine replicators in city and global contexts, and 2) the role of base rate differences on the likelihood of a potential replicators in the population.

Hypothesis

This study hypothesizes that existing forensic signature analysis methods inadequately account for statistical variability across large populations and fail to incorporate the heterogeneity of counterfeiter capabilities, leading to potential misinterpretations in judicial settings.

Statistical Models and Methodology

We employed statistical probability models to evaluate two main contexts:

1. The replication likelihood of a signature in large populations.
2. The influence of base-rate differences on forgery probability among skilled counterfeiters.

2. CONCLUSIONS

Our analysis demonstrates that the expert's confidence, while statistically valid in isolation, does not account for population effects. This calls for a more nuanced presentation of findings to jurors. Finally, the assumptions underlying base-rate neglect significantly weaken the reliability of high-certainty claims. Forensic analysis must incorporate heterogeneity in population capabilities to enhance its credibility.

Context 1: Probability Claims Involve Inherent Likelihood of Error when Applied to Large/Realistic Populations

Handwriting experts often state their level of confidence in two signatures being a match in terms of probability. Per the ASTM Standard E1658 (see McClary, 2006), frequently used terms in this forum are "inconclusive" and "strong probability", the latter of which is often viewed as strong confidence in the decision being made. This set of terms has been used in forensic practice and as a measurement tool in research designs, which have found that these terms do pertain to differences between forensic examiner and layperson error rates in detection (see Kam et al., 2001).

To explore the use of this term, consider a handwriting expert who states that there is a strong probability that two questioned signatures are genuine matches, and that these probabilities are reflected as **99%** and **99.9%,** respectively. This means there is a **1% probability** in one case and a **0.1% probability** in the other that the signatures were made by someone else (a "false match rate"). For a 1% false match rate, this means that, on average, for every **100 people, 1 person** could potentially replicate the signature. For a .1% false match rate, this is equivalent to 1 person for every **1,000**

International Journal for Public Policy, Law and Development
Volume 1, Issue 1, Year 2024

people.

City Population Scenario

We consider a city with a population of **600,000,** where the expected number of potential replicators is:

- **6,000 people** for a signature with **99% authenticity** (600,000 × 0.01 = 6,000).
- **600 people** for a signature with **99.9% authenticity** (600,000 × 0.001 = 600).

Next, we calculate the probability that no one in the city could replicate the signatures.

For the signature with **99% authenticity**, the probability that **no one out of the 600,000 people** replicates it is calculated as:

$$0.99^{600,000} \approx 2.06 \times 10^{-2617}$$

This means the probability that no one could replicate the signature is approximately **0.000...00206%** (with **2,615 zeros** before the 2). This number is extremely close to zero, meaning the chance that no one in the city could replicate the signature is virtually nonexistent. In other words, there is nearly **100% probability** that at least one person in the city could replicate it.

For the signature with **99.9% authenticity**, the probability that no one replicates it is:

$$\mathbf{0.999^{600,000} \approx 1.1963 \times 10^{-261}}$$

This means that the probability of no one replicating the signature is 0.000…001963% **(**with 261 zeros before the 1), which implies that also in this case there is a near **100% chance** that at least one person in the city could replicate the signature.

Global Population Scenario

Let's now expand the analysis to the global population. With the world population estimated at approximately **8 billion people**, the number of potential replicators grows significantly.

For a population of 8 billion:

- **80 million people** could potentially replicate the signature with **99% authenticity** (8,000,000,000 × 0.01 = 80,000,000).
- **8 million people** could potentially replicate the signature with **99.9% authenticity** (8,000,000,000 × 0.001 = 8,000,000).

Next, we calculate the probability that no one in the world population could replicate the signature. For the signature with **99% authenticity**, the probability that **no one out of the 8 billion people** replicates it is:

$$0.99^{8,000,000,000} \approx 2.06 \times 10^{-35,229}$$

This value is so close to zero that it can be considered **virtually zero**. Therefore, the probability that at least one person in the world could replicate the signature is **effectively 100%**. For the signature with **99.9% authenticity**, the probability that **no one replicates it** is:

$$0.999^{8,000,000,000} \approx 4.32 \times 10^{-3,478}$$

International Journal for Public Policy, Law and Development
Volume 1, Issue 1, Year 2024

This is also essentially **zero**, meaning the probability that at least one person in the world could replicate the signature is **virtually 100%**.

Context 1 Conclusion

An expert opinion that states a strong probability (**99%** or **99.9% probability)** that the questioned signatures are genuine, seems, at first glance, to offer high confidence in the authenticity of the signatures. However, upon closer examination and analysis of the probabilities in the context of larger populations, it becomes evident that these probabilities are not as conclusive as they may appear. In such contexts, the probability of false matches rises substantially, and it becomes clear that even a small false match rate leads to a high likelihood of replication by numerous individuals. Therefore, the expert's high certainty does not guarantee that the signature in question could not have been made by someone else—especially when considered against a broader population. This context suggests that confidently declared matches may need to be weighted differently in the eyes of jurors. These probabilities, though seemingly precise, do not offer meaningful assurance of authenticity and should be interpreted with caution.

Context 2: Population Characteristics Alter Probability Base Rates

The second context involved base rates. As often suggested by court presentations, probabilities are affiliated with the likelihoods of replicators in the general population. However, individuals who are not counterfeiters, nonprofessional counterfeiters, and professional counterfeiters are all from distinct populations, who have different capacities in terms of counterfeiting skill. What is the capacity of a professional counterfeiter to effectively counterfeit a signature? This is an interesting empirical question which could be reflected by the base rates in the population. Recent work has suggested that, in the case of handwritten signatures, individuals vary considerably in their capacity as counterfeiters, and that it is relatively easy to train individuals to develop handwriting counterfeiting skills (Ballard et al., 2007). Therefore, when considering a population of **professional counterfeiters**, or even those who have what some would consider adequate counterfeiting capabilities, the analysis of signature replication probabilities becomes even more complex and potentially alarming. Here are some key considerations and thoughts on how the statistical model would shift:

Higher Base Probability of Replication

Professional counterfeiters possess **specialized skills** that allow them to forge signatures with a much higher degree of accuracy than an average person. The **false match rate** of **1% or 0.1%,** which applies to the general population, would likely be **significantly higher** in a population of trained forgers. Even though the base probability reported by the expert (99% or 99.9% certainty) assumes a low likelihood of replication by someone else, this would be an **underestimate** when dealing with professionals. For example, in a population of counterfeiters, it could be assumed that the chance of replicating a signature might not be **1 in 100** or **1 in 1,000** but rather **1 in 10** or even **1 in 5**, depending on their level of skill and expertise.

Increased Number of Potential Replicators

With a higher probability of replication in a population of counterfeiters, the expected number of potential replicators increases dramatically:

- In a population of **10,000 professional counterfeiters**, even if we conservatively assume a **10% false match rate**, **1,000 individuals** could potentially replicate the signature at a level that would fool an expert.
- If the false match rate is **20%**, then **2,000 individuals** could replicate the signature.

This drastically changes the interpretation of expert reports that offer a **99% or 99.9% certainty** because, in this scenario, **many individuals** within a specialized population could replicate the signature. It is also currently unclear if forensic signature experts utilize this possibility in their subjective confidence rating, raising questions as to what exactly the 99%/99.9% probability rating truly represents.

International Journal for Public Policy, Law and Development
Volume 1, Issue 1, Year 2024

Impact on Legal Certainty

In a legal context, the **certainty provided by experts** becomes far less reliable when considering a population of counterfeiters. Even if the signature is given a **99.9% probability of authenticity**, it no longer guarantees that the signature couldn't have been made by someone else. Given the specialized skills of counterfeiters, the certainty provided by the expert could be challenged more effectively. For instance, if even **1 in 10** counterfeiters could replicate the signature, and there are **10,000 counterfeiters** in the world, this results in **1,000 potential replicators**, significantly weakening the expert's claim of certainty. This makes it **virtually certain** that at least one or many individuals, within the counterfeiter population, could successfully replicate the signature.

Context 2 Conclusion

In a population of professional counterfeiters, the statistics provided by experts claiming **99% or 99.9% certainty** become significantly less meaningful. The probability that **at least one counterfeiter** can replicate a signature rises dramatically, leading to greater uncertainty.

Forgeries by trained individuals require a more robust analysis and an increased standard of verification to ensure authenticity. The discussion of Context 2 provides important insight into how base rates are likely not incorporated into forensic expert judgment of probability, despite the fact that many individuals in the average, noncriminal population are likely excluded from the overall population of interest where forgeries/counterfeits may have emerged from. Despite this, incorporation of this concept has not grown into the field in a comprehensive fashion.

3. GENERAL DISCUSSION

The current commentary sought to explore how probabilistic claims made my forensic signature experts may be flawed from a statistical standpoint, thus leading to issues with interpretation and application of forensic expert evidence in court cases. In our commentary, we reviewed to major contexts where these errors and misjudgments may emerge. First, we discussed the number of replicators and false match rates in real-world contexts, showcasing that in city/global contexts, statistical probability reflected in larger, more realistic population sizes does not create a context of high confidence in potential matches. In fact, even low probabilities of error become meaningful when placed in the context of city or global populations. Next, we demonstrated how probable matches are based on probabilities of a uniform/homogenous population when in reality, forgery capacities differ from individual to individual and are likely particularly high in expert counterfeiter subgroups. Given this, a base rate concern arises in presuming that all produces of counterfeits in a city or global population are roughly equivalent in forgery capability.

Context 1 has particular importance for the application of forensic signature evidence in court systems. As noted elsewhere (e.g., Stoels et al., 2012), fallacies about the importance of probabilistic evidence in court can mislead jurors into placing particular weight to evidence which should not be weighed as such. Recent work has found that expert evidence is viewed as quite convincing, especially when the expert is confident and has years of experience (Wilcox & NicDaied, 2018), so evidence not rooted in statistical clarity may still be viewed very positively and convincingly by jurors if delivered in a particular way. As truth is the primary goal of the court systems, this could lead to predictable negative consequences.

Context 2 involves a fallacy which permeates through the final decision of expert evidence, which is that of a base-rate fallacy. Base-rate fallacy (or base-rate neglect) refers to a tendency to ignore base rates when making decisions and, instead, use more intuitive or appealing information to make choices (Pennycook et al., 2022). In the case of forensic expert probability judgments, probability of matches should be made using base rates about the heterogenous counterfeiting capacities which emerge across different subgroups in the populations, or at least the fact that expert counterfeiters likely have higher quality counterfeits than the layperson—some have claimed that probabilities should even be based off of the "active criminal population" (Neumann & Audsmore, 2019), although this clearly has issues of its own.

International Journal for Public Policy, Law and Development
Volume 1, Issue 1, Year 2024

In Forensic Handwriting Identification, Ron Morris (Morris, 2020)'s fifth principle (page 130) acknowledges that it may be impossible to reliably identify simulated handwriting in certain cases. Specifically, if a forger can effectively disguise their own handwriting characteristics and accurately imitate another person's style without leaving distinguishing traces, forensic examiners may struggle to detect that the writing is a forgery. This highlights a limitation in forensic handwriting analysis, as there may not always be clear indicators that writing is a simulation, especially when done skillfully. In practice, while forensic handwriting experts can often detect attempts at imitation through subtle inconsistencies or lack of fluidity, this principle admits that a perfectly executed forgery—where the forger has obscured personal characteristics—might evade detection.

It is important to note that, accuracy aside, this lack of accounting for skill differences in the criminal population may be why jurors question document evidence more frequently than other types of evidence—a recent study found that high probability matches were only viewed as ~65% accurate in the case of handwriting evidence compared to the nearly 90% accuracy perceived in the case of DNA evidence (Ribeiro et al., 2019).

Regardless of juror perception, one might consider amendments to the probability judgment as the marker of forensic evidence. Some suggestions from the literature include automation of document matching using deep learning approaches (Al Neaimi et al., 2020). Other suggestions involve using microscopic analyses of pen pressure to differentiate between signatures (Neto et al., 2021). While technological approaches emerge and become more widespread, perhaps a short-term approach would be to temper the confidence claims made by experts. Even amongst experts, categorical statements remain the predominant manner of communicating findings in court cases, despite the rise in criticism of such statements over recent years (Bali et al., 2020).

The findings highlight systemic issues in forensic signature analysis, particularly regarding statistical validity and communication. Addressing these issues requires a shift towards evidence-based practices, emphasizing transparency and rigorous analysis.

Future work should focus on integrating advanced statistical models and technology to improve forensic reliability and to ensure that forensic signature experts can pass along their judgments to the courts without misrepresenting the conclusions or misleading jurors inadvertently.

Conflict of Interest Statement

The authors have no competing interests to declare

References

1. Al Neaimi, M., Al Hamadi, H., Yeun, C. Y., & Zemerly, M. J. (2020). Digital forensic analysis of files using deep learning. In 2020 3rd International Conference on Signal Processing and Information Security (ICSPIS) (pp. 1-4). IEEE.

2. Ballard, L., Lopresti, D., & Monrose, F. (2007). Forgery quality and its implications for behavioral biometric security. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 37(5), 1107-1118. https://doi.org/10.1109/TSMCB.2007.903539

3. Bali, A. S., Edmond, G., Ballantyne, K. N., Kemp, R. I., & Martire, K. A. (2020). Communicating forensic science opinion: An examination of expert reporting practices. Science & Justice, 60(3), 216-224.

https://doi.org/10.1016/j.scijus.2019.12.005 Crot, S., & Marquis, R. (2024) A comparative review of error rates in forensic handwriting examination. Journal of Forensic Sciences. Advanced Online Publication. https://doi.org/10.1111/1556-4029.15589

4. Hicklin, R. A., Eisenhart, L., Richetelli, N., Miller, M. D., Belcastro, P., Burkes, T. M., ... & Eckenrode, B. A. (2022). Accuracy and reliability of forensic handwriting comparisons. Proceedings of the National Academy of Sciences, 119(32), e2119944119. https://doi.org/10.1073/pnas.2119944119

International Journal for Public Policy, Law and Development
Volume 1, Issue 1, Year 2024

5. Kafadar, K. (2015). Statistical issues in assessing forensic evidence. International Statistical Review, 83(1), 111-134. https://doi.org/10.1111/insr.12069

6. Kam, M., Gummadidala, K., Fielding, G., & Conn, R. (2001). Signature authentication by forensic document examiners. Journal of Forensic Sciences, 46(4), 884-888.

7. Kang, T. Y., Kim, H., Yook, S., & Lee, J. (2022). A study on factors that affect error rates in handwriting examinations of Korean characters by forensic document examiners and non-experts. Forensic Science International, 334, 111266. https://doi.org/10.1016/j.forsciint.2022.111266

8. Kukucka, J., & Kassin, S. M. (2014). Do confessions taint perceptions of handwriting evidence? An empirical test of the forensic confirmation bias. Law and human behavior, 38(3), 256.

9. Lee, G. K., Yap, B. S., Yang, C. Y., Lee, L. T., Tan, S. K., & Tan, K. P. (2006). A study on the levels of difficulty in the simulation of individual characteristics in a signature. Journal of the American Society of Questioned Document Examiners, 9(1).

10. Martire, K. A., Growns, B., & Navarro, D. J. (2018). What do the experts know? Calibration, precision, and the wisdom of crowds among forensic handwriting experts. Psychonomic bulletin & review, 25, 2346-2355. https://doi.org/10.3758/s13423-018-1448-3

11. McClary, C. R. (2006). Conclusions and Guidelines. In Scientific examination of questioned documents (pp. 413-416). CRC Press.

11. Neto, O. N., Sarkis, J. E. S., Siqueira, A. N., Ribeiro, L. A. S., & Freitas, A. Z. (2021). A new method for measuring pen pressure in forensic handwriting analysis–a proof of concept study. Analyst, 146(6), 1973-1980. https://doi.org/10.1039/D0AN01787F

12. Neumann, C., & Ausdemore, M. (2019). Communicating forensic evidence: Is it appropriate to report posterior beliefs when DNA evidence is obtained through a database search? Law, Probability and Risk, 18(1), 25-34. https://doi.org/10.1093/lpr/mgz002

13. Pennycook, G., Newton, C., & Thompson, V. A. (2022). Base-rate neglect. In Cognitive illusions (pp. 44-60). Routledge.

14. Ribeiro, G., Tangen, J. M., & McKimmie, B. M. (2019). Beliefs about error rates and human judgment in forensic science. Forensic science international, 297, 138-147. https://doi.org/10.1016/j.forsciint.2019.01.034

15. USCoAftN, C. DAUBERT et ux., individually and as guardian's ad litem for DAUBERT, Daubert v. Merrell Dow Pharmaceuticals. (1993).

16. Wilcox, A. M., & NicDaeid, N. (2018). Jurors' perceptions of forensic science expert witnesses: Experience, qualifications, testimony style and credibility. Forensic science international, 291, 100-108. https://doi.org/10.1016/j.forsciint.2018.07.030