Original Article

Arab Journal of Forensic Sciences and Forensic Medicine 2024; Volume 6 Special Issue (ASFSFM 2023), 104-116

Naif Arab University for Security Sciences
Arab Journal of Forensic Sciences and Forensic Medicine
المجلة العربية لعلوم الأدلة الجنائية والطب الشرعي
https://journals.nauss.edu.sa/index.php/AJFSFM

Evaluation of GLOCK 9 mm Firing Pin Aperture Shear Mark Individuality Based on 3,156 Different Pistols (Manufactured Over a 30 Year Period in Two Countries) Using Additional Pattern Matching and IBIS Pattern Recognition

تقييم فردية علامة إبرة الإطلاق في مسدسات جلوك عيار 9مم بناءً على الإطلاق من 3156 مسدسًا مختلفًا (صُنعت على مدار 30 عامًا في بلدين) باستخدام مطابقة الأنماط الإضافية والتعرف على الأنماط IBIS

James E. Hamby[1*], Eric Warren[2], Stephen Norris[3], and Nicholas D.K. Petraco[4,5]

[1]International Forensic Science Laboratory & Training Centre, Indianapolis, IN, 46227, United States.
[2]SEP Forensic Consultants, Memphis,TN, United States.
[3]Wyoming State Crime Laboratory, Cheyenne, WY, United States.
[4]Department of Sciences, John Jay College of Criminal Justice, City University of New York, New York, NY, United States.
[5]Faculty of Chemistry and Faculty of Criminal Justice, Graduate Center, City University of New York, New York, NY, United States.

Received 25 Aug. 2023; Accepted 30 May 2024; Available Online 22 July 2024.

Abstract

Over a period of 30 years, a number of fired GLOCK cartridge cases have been evaluated. A total of 3156 GLOCK firearms were used to generate a sample of the same size. Our research hypothesis was that no cartridge cases fired from different 9 mm semiautomatic GLOCK pistols would be mistaken as coming from the same gun (a false match). Using optical comparison microscopy, two separate experiments were carried out to test this hypothesis. A subsample of 617 test-fired cartridge cases were subjected to algorithmic comparison by the Integrated Ballistics Identification System (IBIS). The second experiment subjected the full set of 3,156 cases to manual comparisons using traditional pattern matching. None of the cartridge cases were "matched" by either of these two experiments. Using these empirical findings, an established conservative Bayesian probability model was used to estimate

Keywords: Forensic sciences, Daubert, firearms identification, microscopic examination, false match rate, random match probability, legal challenges, pattern matching, Glock

Production and hosting by NAUSS

المستخلص

على مدار 30 عامًا، تم تقييم عدد من أظرف الطلقات النارية المستخدمة في مسدسات جلوك. وتم استخدام ما مجموعه 3156 سلاحاً نارياً من نوع جلوك لتوليد عينة بنفس الحجم. كانت فرضيتنا البحثية هي أنه لن يتم الخلط بين أي من أظرف الطلقات النارية التي تم إطلاقها من مسدسات جلوك نصف آلية مختلفة عيار 9 مم على أنها قادمة من نفس السلاح (مطابقة خاطئة). وتم إجراء تجربتين منفصلتين لاختبار هذه الفرضية باستخدام مجهر المقارنة البصرية. ثم خضعت عينة فرعية مكونة من 617 حبة قذيفة نارية تم اختبار إطلاقها للمقارنة الخوارزمية بواسطة نظام تحديد المقذوفات المتكامل (IBIS). التجربة الثانية خضعت المجموعة الكاملة المكونة من 3156 طلقة نارية للمقارنات اليدوية باستخدام مطابقة النمط التقليدية. ولم يتم «مطابقة» أي من الطلقات النارية في أي من هاتين التجربتين. باستخدام هذه النتائج التجريبية، تم استخدام نموذج احتمالية بايزي المحافظ لتقدير احتمال أن يتم

الكلمات المفتاحية: علوم الأدلة الجنائية، معيار داوبرت، تحديد السلاح الناري، الفحص المجهري، معدل المطابقة الخاطئة، احتمال المطابقة العشوائية، التحديات القانونية، مطابقة النمط، جلوك.

* Corresponding Author: James E. Hamby,
Email: jimhamby14@aol.com

doi: 10.26735/TEPX3582

the chance that a 9 mm cartridge case, fired from a GLOCK, could be mistaken as coming from the same firearm when in fact it did not (i.e., the false match probability).

الخلط بين أظرف الطلقات النارية عيار 9 مم، تم إطلاقها من مسدس جلوك، على أنها قادمة من نفس السلاح بينما في الواقع لم يكن كذلك (أي احتمال المطابقة الخاطئة).

1. Introduction

Since at least 1925, forensic examiners have used optical comparison microscopy to examine firearm and toolmark evidence [1]. For fired bullets, the trained firearm examiner microscopically evaluates the fine scratches (striae), while for fired cartridge cases, the trained firearms examiner evaluates both the striae and impressions (or “features”) found on bearing surfaces. The examiner then discriminates between features that randomly occur and toolmarks imparted during the manufacturing process [2,3]. Individual features that are determined to occur at random form a pattern unique to a specific firearm. Moran provides excellent information concerning the AFTE Theory of Identification and Range of Conclusions used by the trained firearms examiner community [2]. We also note that firearm and toolmark identifications can always be—and quite often are, as required in most accredited laboratories—verified by another qualified examiner.

In order to test the hypothesis that individual random features found on fired bullets and cartridge cases are unique to the specific firearm that imparted them, extensive empirical research has been conducted and reported on for the past 100+ years. Three excellent references by Ronald Nichols, including a presentation at the ATF Laboratory in California [4–6], comprise a very comprehensive review of the literature that pertains to firearm and toolmark identification criteria.

In 2011, Petraco, a senior forensic scientist retired from the New York Police Department Crime Laboratory, authored a textbook covering the collection and examination of impression evidence based on AFTE training materials. This reference book has been well received in the forensic and academic communities [7,8]. Additional articles by Grzybowski et al. and Biasotti et al. [9–11] offer a valuable compendium of reference materials that discuss scientific methods and the reliability and validity of the field of firearm and toolmark identification. For additional discussion of research that has been conducted for cartridge cases through 2013, see previous work [12].

In 2014, a landmark study was performed by Iowa State University. This was the first “black box” study performed in the field of firearms identification. The researchers test fired 25 new Ruger SR9 firearms and prepared tests consisting of 15 sets of 3 known test fires with 1 unknown cartridge case. Participants were asked how many of the 3 knowns were suitable for comparison and whether or not the unknown was fired in the same firearm as the knowns in that particular set [13].

In 2014 and 2015, several groups published studies examining cartridge cases using 3D imaging. Both consecutively [14] and non-consecutively [15,16] manufactured firearms were used to fire cartridge cases. Even in the worst-case-scenario of consecutively manufactured firearms, algorithms on the 3D images could correctly identify the images to the correct firearm, further illustrating that tools impart unique surface topographies which are identifiable.

2016 gave rise to a presidential report known as the PCAST report [17] that concluded black box studies like the 2014 Baldwin study, known as the Ames I study, [13] were the only acceptable type of study to establish the scientific validity of a field. PCAST defined a black box study as an empirical study that assesses the opinions reached from a comparison. This would be opposed to a white box study that assesses how opinions were reached in a comparison. They further concluded that two black box studies would be sufficient to address this question. Although firearms identification only had one at the time the PCAST report was published, others soon followed.

A white box study was also published in 2016 by Smith, et al. which was designed to look at factors involved in the decision making process as well as calculate the classical error rate for the participants. Tests were designed with 12 bullets and 12 cartridge cases in an open-set design. The false-positive error rate (FPR) was 0.144% and 0.000% and the false-negative error rate (FNR) was 0.433% and 0.105%, for cartridge cases and bullets respectively. Additionally, the overall error rate for this study was reported to be 0.303% [18].

In 2017, a NIST study was published that demonstrated the use of the Congruent Matching Cell (CMC) algorithm to compare 3D topographical measurements of cartridge cases sourced from a previously distributed proficiency tests. The objective computer algorithm utilized by the authors reveals quantitatively what examiners observe and report on qualitatively; leading the authors to conclude that it is "possible to relate qualitative similarity to a quantifiable similarity metric." [19]. Although not acknowledged by the authors, this research also further supports the idea that inconclusive decisions are necessary and appropriate in firearms examinations due to the overlap observed in some of the histograms where the algorithm struggled to accurately make a binary choice.

Song et al. in 2018 further refined their CMC algorithm and measured the error rate of the algorithm using topographical data from 40 cartridge cases fired in 10 consecutively manufactured Ruger 9 mm slides and 95 cartridge cases fired in 11 non-consecutively manufactured Ruger 9 mm slides. Using the forced binary choice design, the authors calculated FPRs and FNRs for each dataset. The FPR and FNR for the consecutively manufactured dataset was 5.6 x 10-19 and 1.2 x 10-3, respectively; while the FPR and FNR for the non-consecutively manufactured dataset was 1.8 x 10-17 and 2.2 x 10-4, respectively [20].

Also in 2018, a second black box study was published by Keisler, et al. In this study, the authors created test kits with 20 sets of two cartridge cases each. Participants were asked to evaluate each set and make a decision of identification, inconclusive, or exclusion. The authors of this study reported the overall error rate to be 0%. Furthermore, the sensitivity was calculated as 99.7%, and the specificity was 79.9% [21]. The successful completion of this second black box study was deemed to meet the earlier PCAST criteria for foundational validity.

In a study in 2019 [22], researchers applied 3D topographical similarity scoring criteria to bullets from 3 different sets of a previous study known as the “Ruger 10 barrel test” [23] as well as a proprietary set of 406 bullets from 136 various firearms. The authors found similar error rates as those seen previously for cartridge cases (between 1 x 10-6 and 1 x 10-10], indicating that the objective comparison of bullets and cartridge cases by computer algorithms support and complement visual comparisons by properly trained human examiners. The same year, two of the authors of this study published an update on the previous “Ruger 10 barrel test”. They examined the error rate of 697 participants from 32 different countries, using both optical comparison microscopy as well as 8 different computerized imaging systems. Based on their analysis, the error rate for this test was determined to be 0.053% with a 95% probability interval of [1.1 x 10-5%, 0.16%] [24].

In 2020, Law and Morris used the CMC algorithm to examine the similarity between cartridge cases and double-casts of those same cartridge cases in order to determine how well the double-casting process replicates the surface topography of a cartridge case. This study determined that double-casting faithfully reproduces the surface topography of cartridge cases. Due to the demonstrated ability to provide participants with virtually identical copies of the same samples, this method provides a method for consistent training across examiners and better reproducibility in proficiency tests as well as validation studies [25].

Work published by the Center for Statistics and Applications in Forensic Evidence at Iowa State University (CSAFE) in 2020 also applied different similarity scoring algorithms to bullets. In this study, the authors found significant ambiguity in the SAM CMS scoring due to same-source distribution overlap with different-source distribution, concluding that a binary decision threshold or cutoff value does not exist that completely separates same-source scores from different-source scores across all three scoring measures [26]. The presence of a third category would potentially resolve this discrepancy and further supports the appropriateness for an inconclusive category as reflected in actual casework.

In 2021, Chapnick et al. published a study using Virtual Comparison Microscopy (VCM). Their study involved 107 participants with each participant examining 40 test sets of fired cartridge cases from firearms with a variety of makes, models, and calibers. By using VCM, the participants all viewed the same items and were asked to annotate the compared sets. This allowed a classical error rate to be calculated but also provided heat map illustrations of the areas found to be similar or different by participants. The overall FPR and FNR were 0.43% and 0.00%, respectively [27].

In 2021 another study was conducted by Law and Morris [28]. In this open set study, double casts of cartridge cases were created and sent to study participants for examination and comparison. The examiners who participated did not have the firearm used to generate the original cartridge cases. This study had a small sample size of 18 examiners who volunteered to participate in the study. Also, while choosing to include "inconclusive" as a response, the authors only considered one subcategory of inconclusive described in the AFTE range of conclusions to be a "true" inconclusive. Depending on the ground truth, inconclusive conclusions of the other two categories were considered incorrect responses by the authors, though they are a normal part of the standard of practice. The participants of the study do not seem to have been made aware of the authors study design decision to mark two subcategories of inconclusive conclusions contained in the AFTE range of conclusions as incorrect. The study found accuracy issues and variability within the conclusions given by the examiners, however as noted by the authors the accuracy/variation issues were almost exclusively contained within the inconclusive categories. One false positive conclusion was observed.

In 2022, Monson et al. described a study they conducted (known as Ames II) and the factors for which they accounted in the design of the study. In this study, which was a double-blind, open set, black box study, they produced test kits consisting of 30 comparison sets. Of these sets, 15 were 2 known cartridge cases and 1 unknown cartridge case to be compared to the knowns in that set. The remaining 15 were of the same design but with bullets instead of cartridge cases. The items in these sets were produced from 23 consecutively manufactured Beretta slides and barrels, 10 Jimenez consecutively finished slides and barrels, and 4 random Beretta slides and barrels from the FBI Reference Firearms Collection (RFC). The authors stated that "[a]nalysis of experimental results will be presented in a series of forthcoming publications" [29]. Indeed, later the same year, the authors did publish results which revealed a FPR of 0.656% and 0.933% for bullets and cartridge cases, respectively and a FNR of 2.87% and 1.87% for bullets and cartridge cases, respectively. The authors note that "[t]he majority of errors were made by a limited number of examiners" and further that their results "are consistent with prior studies, despite its comprehensive design and challenging specimens" [30].

In this study, the authors continue to explore the likelihood of whether or not cartridge cases fired from different model GLOCK semiautomatic 9 mm pistols might be incorrectly matched to the wrong firearm by a qualified firearms examiner or the IBIS computer aided identification system. That is, the operating hypothesis of this study is that no cartridge cases fired from different 9 mm semi-automatic GLOCK pistols should be determined to match each other, by either human comparison using pattern recognition (a trained firearms examiner) or machine comparison (IBIS). GLOCKs are ideal in the sense that they are commonly encountered in case work and are very well known to generally produce well-defined firing pin aperture shear on the primer of cartridge cases fired in them. Thus, a false match rate estimate on these could provide a "baseline" lower bound on the false match rate for more difficult toolmark comparisons.

Given the empirical findings detailed below, an established Bayesian model was then employed to estimate the probability of falsely matching an expended 9 mm cartridge case from a semiautomatic pistol that did not fire it [31]. Such an error rate estimate is the type of "quality assurance" embodied in the Daubert standard.

2. Materials and Methods

As previously described, over a one-month period in 1997, 617 9 mm GLOCK pistols were each test fired four times to obtain fired cartridge cases for evaluation, which were then collected into envelopes with randomly assigned control numbers from 1 to 617. The cartridge cases were subsequently removed from each envelope, scribed with the assigned control numbers, and placed in plastic ammunition trays for examination. Initially, all four test fires were intercompared using optical comparison microscopy as a control to ensure the individual characteristics present were reproducing on test fires, then one representative test fired cartridge case from each of the 617 guns were imaged into the Integrated Ballistics Identification System (IBIS) by Specialist John Brooks and a request was made to Forensic Technology, Inc. (FTI), Montreal, Canada, to perform correlations to determine whether any of the cases "matched" to each other (i.e., were misidentified to one another) based solely on the IBIS correlation score. These examinations involved a total of (617 x 616)/2 = 190,036 pairwise comparisons by the IBIS correlation algorithm. As these cartridge cases were all fired from different 9 mm firearms, none of these cartridge cases should be identified to another cartridge case from this study.

A second microscopic examination of the firing pin aperture shear marks by trained firearms examiners was then carried out on this original set of 617 cases, as well as two more groups of cartridge cases collected over the subsequent two decades. The first of these groups consisted of two test fired 9 mm cartridge cases obtained from 700 different GLOCK firearms over a 5-year period as part of their quality assurance program (1,400 cartridge cases). The second of these groups consisted of 315 recently fired GLOCK cartridge cases—including 12 from consecutively manufactured slides obtained by one of the authors from GLOCK. The third group of 1,839 GLOCK cartridge cases was obtained from various forensic laboratories (four test fires from each firearm that were collected to be used as reference samples in the case of an officer involved shooting) and archives from GLOCK (two test fires each, as generated in the factory and examined for quality control). These cartridge cases were aggregated with the original sample of 617 from 1997 and collectively examined using a Leica Model K-2700 comparison microscope (Leica Microsystems, Wetzlar, Germany), an American Optical Model K-1453 Forensic Comparison Microscope, or a Leica Model DMC Forensic Comparison Microscope. Table 1 lists information for the 9 mm ammunition used in this research.

The examination protocol was as follows: the first cartridge case was designated as the primary case and placed on the left side of the comparison microscope. Using the right side of the comparison microscope, the remaining 3,155 fired cartridge cases were compared to the primary cartridge case until all 3,156 cases were examined. After case number 1 was examined against the other cartridge cases, the entire process started again. Cartridge case 2 was then examined against the other 3,154 and so forth until all cartridge cases were examined against each other. This resulted in (3,156 x 3,155)/2 = 4,978,590 unique pairwise comparisons by a trained human forensic firearms examiner (partially by Dr. Hamby, and partially by co-author Steve Norris). For reference, Figures 1 and 2 show examples of both a "nonmatch" and a "match" for 9 mm GLOCK pistol firing pin aperture shear marks.

Although not included in this analysis, it is im-

Table 1- Types of 9 mm ammunition evaluated during this research project

Headstamp	Case	Primer
*FC 10	Brass	Silver
Tulammo	Steel	Bronze
FC NT	Brass	Bronze
FC	Nickel	Silver
CCI	Aluminium	Silver
* WCC 04	Brass	Brass
* WCC 05	Brass	Brass
R - P	Brass	Brass
R - P	Brass	Silver
WIN	Brass	Silver
Barnes	Nickel	Silver
Speer	Brass	Silver
Hornady	Nickel	Silver
Pro Grade	Brass	Silver
GECO	Brass	Silver
WIN	Brass	Brass
Hornady	Nickel	Brass
RTAC	Brass	Silver
Speer 13	Nickel	Silver
Aguila	Brass	Brass
GFI	Brass	Silver
CBC	Brass	Brass

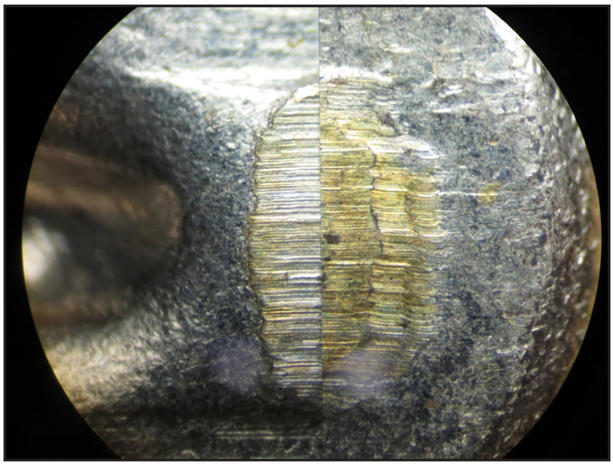

Figure 1 - *Example of a 9 mm GLOCK firing pin aperture shear mark known non-match (KNM)*

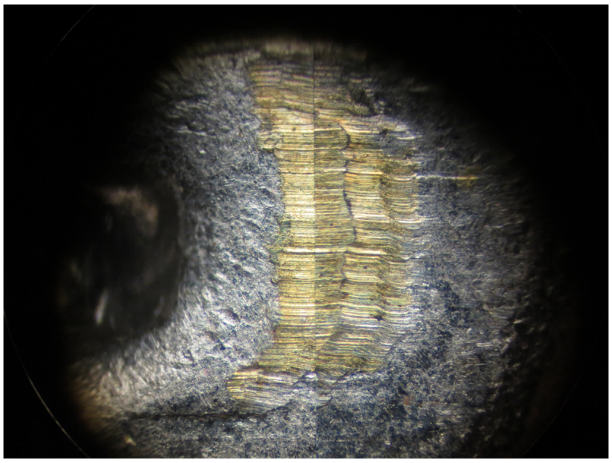

Figure 2 - *-Example of a 9 mm GLOCK firing pin aperture shear mark known match (KM)*

portant to note that in addition to examination by trained firearms examiners, a subset of these cartridge cases was utilized in a firearms examiner training program developed by the International Forensic Science Laboratory & Training Centre, where it was used to train new firearms examiners in Belize, Botswana, Jamaica, Guam, and several US crime laboratories. In these exercises, trainees were presented with different combinations of same source and different source comparisons. Types of exercises such as these are regularly used as training tests to help new firearms examiners to develop their threshold for source attribution conclusions or competency tests to complete training. Furthermore, these specific descriptive assessment training exercises serve as somewhat of a blind verification of the initial findings in this study.

3. Statistical Analysis

Based on the stated hypothesis, this study was designed to estimate the chances of a “false match”, or that the firing pin aperture shear marks on cartridge cases fired from two different 9 mm GLOCKs would be assessed as having been fired from the same 9 mm GLOCK. From a computational pattern recognition point of view, the human examiner or the IBIS is playing the role of a “classifier” [32]. Thus, rephrasing slightly, we wish to estimate the “classifiers” expected false match rate (FMR). We would like to have a measure of uncertainty in our estimate with some given level of probability (i.e., an interval around the expected FMR) as well.

Typically, in a study of this design, a frequentist-based approach is used to estimate the FMR which exploits the binomial distribution [22,33]. This is problematic for two main reasons: First, no false matches were observed in this study (discussed below) and thus baring ad hoc techniques, no interval estimates for FMR are possible. Second, when multiple firearms are tested multiple times, the binomial model can underestimate variability [33,34]. This is true not only in our case, but in any design using pairwise comparisons [32].

This is where Bayesian methodology can be of some assistance. All unknowns are treated as random variables in Bayesian statistics. “Knowledge” in quantities of interest is represented as a probability distribution. A Bayesian technique takes what is “known” or “believed” about an unknown parameter (FMR in our case) and represents it as a prior distribution. Then, updating this with data via a likelihood function, one obtains a posterior distribution on the parameter(s) of interest. Everything we currently “know” about the parameter (FMR for this study) in light of data from the experiment is summarized by its posterior distribution. Bayes theorem can be compactly expressed in our situation as:

$$p(\mathrm{FMR}|\mathrm{data}) = \frac{p(\mathrm{data}|\mathrm{FMR})}{p(\mathrm{data})} p(\mathrm{FMR})$$

On the left-hand side is the posterior probability density for FMR. The “data” in our case are the results of the human/IBIS comparisons. The rightmost term is the prior probability density for our belief as to what the FMR is before observing the experimental data. Here, we would like to assume little and take the prior to be fairly uniform on the interval (0,1). The term in the denominator is often referred to as the “evidence” and is difficult to compute directly. However, with the technique we will employ for this study (Markov chain Monte Carlo sampling), it can be ignored. The remaining term on the right-hand side is the data’s likelihood function. Schuckers has noted that for this type of study design, a beta-binomial model for the likelihood function can model the data well and incorporate extra variation not accounted for in the binomial model which assumes statistically independent comparisons [31,32]. Couched in Bayes’ theorem, this framework also allows us to easily give an interval estimate of uncertainty around the posterior FMR [31].

Specifically, let n be the number of comparisons the examiner (human or machine) conducted and x the number of false matches made. Table 2 lists n and x for each of the two experiments.

Table 2- Results of comparison experiments.

	IBIS Comparisons[a]	**Human Examiner Comparisonsa**
Sample size	617	3,156
nb	190,036	4,978,590
xc	0	0

For each comparison the examiner (i.e. the classifier) executes is a possibly correlated Bernoulli trial with a probability of a false match p. The number of false matches, x, is modeled as a binomial probability mass function:

$$x \sim \text{binomial}(n, p) \quad \text{i.e.} \;\; \text{binomial}(x|n,p) = \binom{n}{x} p^x (1-p)^{n-x}.$$

The FMR, p, is represented as a beta probability density function with parameters α and β:

$$p \sim \text{beta}(\alpha, \beta) \quad \text{i.e.} \;\; \text{beta}(p|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}.$$

The joint probability density over *x* and *p* is then:

$$f(x,p) = \text{beta}(p|\alpha,\beta) \times \text{binomial}(x|n,p).$$

By integrating over *p*:

$$\int f(x,p)\, dp = \text{beta-binomial}(x|\alpha,\beta,n) = \binom{n}{x} \frac{\Gamma(\alpha+\beta)\Gamma(\alpha+x)\Gamma(\beta+n+x)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha+\beta+n)}$$

we are left with a likelihood function for the data, χ (i.e. a beta-binomial likelihood) which naturally incorporates increased variance over the simpler binomial likelihood model. The increased variance over the binomial likelihood accounts for correlation induced by pair-wise comparisons [31].

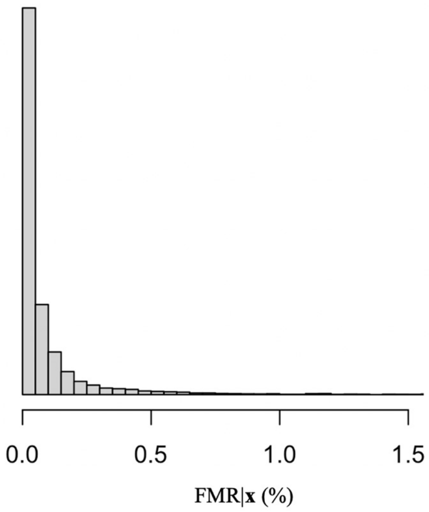

Figure 3 - *-Posterior probability distribution (zoomed in) for false match rate of firing pin aperture shear marks given data obtained in this study.*

At this point we seem to have replaced what we were really interested in, the FMR (i.e. *p*), with α and β. However, FMR can be recovered from α and β as:

$$\text{FMR} = \frac{\alpha}{\alpha+\beta}$$

Half-Cauchy priors with location 0 and scale 100 are assigned for parameters a and b. These priors were chosen because the prior on FMR was fairly evenly spread between 0% and 100% and they implied a moderate amount of correlation between pair-wise comparisons, a priori. The general Hamiltonian Monte-Carlo software Stan [35] was used to sample the posterior for α and β:

$$p(\alpha,\beta|\mathbf{x},\mathbf{n}) \propto \text{beta-binomial}(\mathbf{x}|\alpha,\beta,\mathbf{n}) \times \text{half-Cauchy}(\alpha|0,100) \times \text{half-Cauchy}(\beta|0,100)$$

Eight chains were used with 20,000 iterations each. The first 10,000 iterations of each chain were treated as warm-ups and thrown away. The final chains were thinned to retain only every tenth sample. R-hat convergence diagnostics were all 1.0 (the chains are effectively converged) [36]. Autocorrelation plots showed no sign of significant within chain correlation. A total of 8,000 (marginal) samples for a and b were drawn from the posterior. This posterior sample of a and b was then used to compute the FMR posterior by the equation above.

4. Results and Discussion

The original set of 617 fired cases were digitally imaged and entered into the Integrated Ballistics Identification System (IBIS) for correlation against each other. It was found that the instrument had the capacity and capability to handle the number of cartridge cases and to achieve zero misidentifications. Mike McLean from FTI provided the following information concerning their evaluation. He stated "…Forensic Technology (FTI) was provided test results from 617 test fired exhibits from 617 different GLOCK pistols. FTI then imaged and correlated all of these samples against one another in order to see if any matches were among these samples. It was found that none of the 617 Test Samples were matches to one another…" (M. McLean, personnel communication, June 1992).

Regarding the optical microscopy comparisons, each of the 3,156 fired cartridge cases examined by a human could be seen to have unique individual characteristics. The cartridge cases were therefore individualized to the exclusion of the remaining cartridge cases. Thus, using pattern matching identification criteria with optical comparison microscopy, trained firearms examiners were able to determine that each cartridge case had firing pin aperture striae that would not be mistakenly associated with the wrong firearm. Figure 3 also demonstrates this finding from a statistical perspective. The figure is the (approximate) posterior distribution for the FMR given the data contained in Table 2. It shows the (posterior) probability of a false match rate to have a mean value of 0.1% and a median value 0.03%. The 95% credibility interval for the FMR is [4×10-7%, 0.4%]. Given the false match rate model from these results, it can be inferred that the probability of two different 9 mm semiautomatic Glocks producing firing pin aperture shear marks that could be mistaken as coming from the same Glock is highly unlikely. One limitation of our study is that it is a closed set test design. Closed set designs have been proposed to underestimate the error rate of a study [29, 37].

One of the firearms examiners that performed these examinations had approximately four years of experience performing independent casework when he examined a portion of the cartridge cases in this study. The other firearm examiner had been performing independent casework for approximately 45 years. The firearms examiner trainees were provided selected items from this study approximately one year into their respective training programs. It is noteworthy that recent studies have shown no significant difference in discrimination ability between properly trained examiners with varying years of experience [30, 38].

5. Conclusion

Forensic firearms' examiners have routinely identified or excluded fired bullets and cartridge cases with suspect firearms over the past 100+ years. This research project, involving the comparison of 3,156 fired cartridge cases against each other, empirically validated the premise that each was identifiable and unique. An upper limit to a false match probability for this group of firing pin aperture shear marks on cartridge cases was conservatively estimated to be 0.4% based on these findings. Although no firearm-to-cartridge case identifications were made in this research, the results support the concept of identifying fired cartridge cases to the firearm that fired. A future avenue for further study could include having some of these cartridge cases re-examined blindly as well as subjecting some of these items to 3D surface topography measurements and subsequent virtual comparison microscopy.

Conflict of interest

The authors declare no conflicts of interest.

Source of funding

The authors received no financial support for the research, authorship or publication of this paper.

Acknowledgements

The authors are grateful to everyone that assisted with the test firing, collection, and data entry of the fired cartridge cases. The authors are also grateful to Dr. Peter Diaczuk for providing the images used in Figures 1 and 2.

References

1. Berg SO. The History of Firearms Identification. Identif News. 1965 Jun;15(6):5–15.
2. Moran B. A Report on the AFTE Theory of Identification and Range of Conclusions for Tool Mark Identification and Resulting Approaches to Casework. AFTE J. 2002;34(2):227–35.
3. SWGGUN. Scientific Working Group for Firearms and Toolmarks [Internet]. 2014. Available from: https://web.archive.org/web/20140209075636/http://www.swggun.org/swg/index.php
4. Nichols RG. Firearm and toolmark identification criteria: a review of the literature. J Forensic Sci. 1997;42(3):466–74.
5. Nichols RG. Firearm and toolmark identification criteria: a review of the literature, part II. J Forensic Sci. 2003 Mar;48(2):318–27.
6. Nichols RG. The History of Firearms and Tool Mark Identification Criteria. Proceedings of the 55th Annual Meeting of the American Academy of Forensic Sciences; 2003 Feb 17; Chicago, IL.
7. Petraco N. Color Atlas of Forensic Toolmark Identification. Boca Raton: CRC Press; 2010. 177 p.
8. Dutton G. Book review: “Color Atlas of Forensic Toolmark Identification” By Nicholas Petraco. AFTE J. 2011;43(3):201–3.
9. Grzybowski RA, Miller J, Moran B, Nichols RG, Thompson R. Firearm/Toolmark Identification: Passing the Reliability Test Under Federal and State Evidentiary Standards. AFTE J. 2003;35(2):209–41.
10. Biasotti AA, Murdock JE, Moran B. Chapter 35 - Firearms and Toolmark Identification. In: Faigman DL, Blumenthal JA, Sanders J, Chen EK, Mnookin JK, Murphy EE, editors. Modern Scientific Evidence the Law and Science of Expert Testimony. 2011th–2012th ed. Eagan, MN: West / Thomson Reuters; 2012. p. 641–730.
11. Grzybowski RA, Murdock JE. Firearm and Toolmark Identifcation- Meeting the Daubert Challenge. AFTE J. 1998;30(1):3–14.
12. Hamby JE, Norris S, Petraco NDK. Evaluation of GLOCK 9mm Firing Pin Aperture Shear Mark Individuality Based On 1,632 Different Pistols by Traditional Pattern Matching and IBIS Pattern Recognition. J Forensic Sci. 2016 Jan;61(1):170–6.

13. Baldwin DP, Bajic SJ, Morris MD, Zamzow D. A Study of False-Positive and False-Negative Error Rates in Cartridge Case Comparisons. Ames, Iowa: Iowa State University; 2014 Apr p. 36. Report No.: NCJ 249874.
14. Zheng XA, Soons JA, Vorburger TV, Song JF, Renegar TB, Thompson RM. Applications of surface metrology in firearm identification. Surf Topogr Metrol Prop. 2014;2(1):014012.
15. Riva F, Champod C. Automatic Comparison and Evaluation of Impressions Left by a Firearm on Fired Cartridge Cases. J Forensic Sci. 2014 May 1;59(3):637–47.
16. Weller T, Brubaker M, Duez P, Lilien R. Introduction and Initial Evaluation of a Novel Three-Dimensional Imaging and Analysis System for Firearm Forensics. AFTE J. 2015;47(4):198–208.
17. President's Council of Advisors on Science and Technology. Report on Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods. Washington, D.C.: Executive Office of the President; 2016 Sep p. 147.
18. Smith TP, Andrew Smith G, Snipes JB. A Validation Study of Bullet and Cartridge Case Comparisons Using Samples Representative of Actual Casework. J Forensic Sci. 2016 Jul 1;61(4):939–46.
19. Ott D, Thompson R, Song J. Applying 3D measurements and computer matching algorithms to two firearm examination proficiency tests. Forensic Sci Int. 2017 Feb 1;271:98–106.
20. Song J, Vorburger TV, Chu W, Yen J, Soons JA, Ott DB, et al. Estimating error rates for firearm evidence identifications in forensic science. Forensic Sci Int. 2018 Mar 1;284:15–32.
21. Keisler MA, Hartman S, Kilmaon A, Oberg M, Templeton M. Isolated Pairs Research Study. AFTE J. 2018;50(1):56–8.
22. Roberge D, Beauchamp A, Lévesque S. Objective Identification of Bullets Based on 3D Pattern Matching and Line Counting Scores. Int J Pattern Recognit Artif Intell. 2019 Feb 19;33(11):1940021.
23. Hamby JE, Brundage DJ, Thorpe JW. The Identification of Bullets Fired from 10 Consecutively Rifled 9mm Ruger Pistol Barrels: A Research Project Involving 507 Participants from 20 Countries. AFTE J. 2009;41(2):99–110.
24. Hamby JE, Brundage DJ, Petraco NDK, Thorpe JW. A Worldwide Study of Bullets Fired From 10 Consecutively Rifled 9MM RUGER Pistol Barrels—Analysis of Examiner Error Rate. J Forensic Sci. 2019;64(2):551–7.
25. Law EF, Morris KB. Three-Dimensional Analysis of Cartridge Case Double-Casts. J Forensic Sci. 2020;65(6):1945–53.
26. Vanderplas S, Nally M, Klep T, Cadevall C, Hofmann H. Comparison of three similarity scores for bullet LEA matching. Forensic Sci Int. 2020 Mar 1;308:110167.
27. Chapnick C, Weller TJ, Duez P, Meschke E, Marshall J, Lilien R. Results of the 3D Virtual Comparison Microscopy Error Rate (VCMER) Study for firearm forensics. J Forensic Sci. 2021;66(2):557–70.
28. Law EF, Morris KB. Evaluating firearm examiner conclusion variability using cartridge case reproductions. J Forensic Sci. 2021;66(5):1704–20.
29. Monson KL, Smith ED, Bajic SJ. Planning, design and logistics of a decision analysis study: The FBI/Ames study involving forensic firearms examiners. Forensic Sci Int Synergy. 2022 Jan 1;4:100221.

30. Monson KL, Smith ED, Peters EM. Accuracy of comparison decisions by forensic firearms examiners. J Forensic Sci. 2023;68(1):86–100.
31. Schuckers ME. Interval estimates when no failures are observed. In: AutoID’02 Proceedings: Workshop on Automatic Identification Advanced Technologies [Internet]. Tarrytown, NY: IEEE; 2002. p. 37–41. Available from: http://myslu.stlawu.edu/~msch/biometrics/papers/autoidpaper.pdf
32. Schuckers ME. Computational Methods in Biometric Authentication: Statistical Methods for Performance Evaluation. 1st ed. New York, NY: Springer Science & Business Media; 2010. 330 p.
33. Wayman JL. Confidence interval and test size estimation for biometric data. San Jose, CA; 2000 Aug. Report No.: MDA904-97-C–03.
34. Agresti A. Categorical data analysis. 3rd ed. Hoboken, N.J: Wiley; 2013. 752 p.
35. Stan Development Team. Stan: A C++ Library for Probability and Sampling [Internet]. 2018. Available from: https://mc-stan.org/
36. Brooks SP, Gelman A. General Methods for Monitoring Convergence of Iterative Simulations. J Comput Graph Stat. 1998 Dec 1;7(4):434–55.
37. Hofmann H, Carriquiry A, Vanderplas S. Treatment of inconclusives in the AFTE range of conclusions. Law Probab Risk. 2020 Dec 1;19(3–4):317–64.
38. Fadul TG, Hernandez GA, Stoiloff SL, Gulati S. An Empirical Study to Improve the Scientific Foundation of Forensic Firearm and Tool Mark Identification Utilizing 10 Consecutively Manufactured Slides. AFTE J. 2013;45(4):376–93.