JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, May 2017, Vol. 62, No. 3
doi: 10.1111/1556-4029.13386
Available online at: onlinelibrary.wiley.com

PAPER

CRIMINALISTICS

John E. Murdock,[1] M.S.; Nicholas D.K. Petraco,[2] Ph.D.; John I. Thornton,[3] D.Crim.; Michael T. Neel,[4] M.S.; Todd J. Weller,[5] M.S.; Robert M. Thompson,[6] B.S.; James E. Hamby,[7] Ph.D.; and Eric R. Collins,[1] B.S.

The Development and Application of Random Match Probabilities to Firearm and Toolmark Identification*

ABSTRACT: The field of firearms and toolmark analysis has encountered deep scrutiny of late, stemming from a handful of voices, primarily in the law and statistical communities. While strong scrutiny is a healthy and necessary part of any scientific endeavor, much of the current criticism leveled at firearm and toolmark analysis is, at best, misinformed and, at worst, punditry. One of the most persistent criticisms stems from the view that as the field lacks quantified random match probability data (or at least a firm statistical model) with which to calculate the probability of a false match, all expert testimony concerning firearm and toolmark identification or source attribution is unreliable and should be ruled inadmissible. However, this critique does not stem from the hard work of actually obtaining data and performing the scientific research required to support or reject current findings in the literature. Although there are sound reasons (described herein) why there is currently no unifying probabilistic model for the comparison of striated and impressed toolmarks as there is in the field of forensic DNA profiling, much statistical research has been, and continues to be, done to aid the criminal justice system. This research has thus far shown that error rate estimates for the field are very low, especially when compared to other forms of judicial error. The first purpose of this paper is to point out the logical fallacies in the arguments of a small group of pundits, who advocate a particular viewpoint but cloak it as fact and research. The second purpose is to give a balanced review of the literature regarding random match probability models and statistical applications that have been carried out in forensic firearm and toolmark analysis.

KEYWORDS: forensic science, coincidental match probability, correspondence probabilities, Daubert, empirical error rate, error rate, false match error, firearm and toolmark identification, likelihood ratio, likelihood ratio, probability, random match probabilities, research, random match probability, statistics, uncertainty

In an adversarial system of justice, the form, substance, and reliability of evidence will always be subjected to scrutiny. Under the premise that the validity of evidence offered to establish relevant facts will be graced and hopefully enhanced by a full and perhaps contentious airing of that evidence, such scrutiny is not only permitted, but encouraged. In this manner, improper interpretation and substandard work may be identified and rejected. But, just as substandard work deserves derision, so does substandard criticism.

Severe criticism of the physical evidence area of firearm and toolmark identification has recently been levied by a number of legal academics. These criticisms have taken several forms. One is that firearm and toolmark identification is highly subjective and therefore is suspect on that basis alone. Another criticism is related to the uniqueness of firearm and toolmark evidence and issues related to its expression. Yet another criticism is that the community of firearm and toolmark examiners has failed to supply the "known or potential error rate" advocated as one of the potential considerations in the Daubert decision (1). Of these three primary criticisms, the last one is the most difficult to address. Schwartz states emphatically that "If no firm statistical basis for firearms and toolmark identification has been demonstrated, there is no scientific basis for any testimony about firearm and toolmark matches" (2; emphasis from original text). Carriquiry opines that unless the field of firearm and toolmark identification has coincidental match probability estimates (i.e., random match probability, or RMP, one type of statistical assessment), the probative value of the evidence cannot be evaluated (3). All of this is little more than academic sophistry and at least one court has noted that this viewpoint is not informed by actual research (4). Schwartz and Carriquiry ignore the inescapable fact that firearm and toolmark identification works, and to categorically reject firearm and toolmark evidence would be akin to saying that the risk of deaths from automobile accidents is sufficiently high enough to justify banning everyone from driving automobiles. While discussion of this topic is reasonable, the discussion must be projected fairly against a solid fact:

---

[1] Contra Costa County Office of the Sheriff, Forensic Services Division, Martinez, CA.
[2] John Jay College of Criminal Justice, CUNY, Manhattan, NY.
[3] Independent Forensic Scientist (Retired), Napa, CA.
[4] Bureau of Alcohol Tobacco, Firearms and Explosives Forensic Laboratory, Atlanta, GA.
[5] Oakland Police Department Crime Laboratory, Oakland, CA.
[6] National Institute of Standards and Technology (NIST), Gaithersburg, MD.
[7] International Forensic Science Laboratory & Training Centre, Indianapolis, IN.

*All comments are those of the authors and do not necessarily reflect the opinions or beliefs of their respective agencies.

Received 17 Feb. 2016; and in revised form 26 June 2016; accepted 17 July 2016.

firearm and toolmark identification demonstrably produces accurate conclusions—a point that cannot be dismissed in critical and ethically-reasoned arguments.

Let us closely examine this last criticism, and how it is expressed. Critics have essentially stated that as firearm and toolmark identification does not have a firm statistical basis, it is therefore worthless, as is any testimony concerning it. This rephrasing is not unfair and allows us to assess this criticism more cogently. This attempts to make the criticism more definite, and as Thomas Jefferson put it, "ideas must be made definite before they can be acted upon by reason."

First, it should be recognized that this criticism is an argument, not a statement of fact. It takes a premise—that firearm and toolmark identification lacks an overarching statistical basis—a premise that is accepted by all, and melds that premise with a conclusion—that it is therefore untrustworthy. Furthermore, it is not even a well-grounded argument; it is defective from several standpoints of logic. One, and perhaps the most salient, is that it is *ignoratio elenchi*—an argument that does not address the issue in question, which is whether or not firearm and toolmark identification actually works. This lapse of logic is also known as an "irrelevant conclusion," or more commonly as "missing the point." Another is that it is an *argumentum ad ignorantium*—that is, an assumption that a claim is true when it has not, or cannot, be proven true. And finally, this argument is a straw man—an argument based on a misrepresentation of an opponent's position. The criticism is based on the assumption that firearm and toolmark examiners will report and subsequently testify to an absolute certainty in their findings as opposed to a practical certainty, but in reality the latter is actually the professional norm.

Any notion that the community of firearm and toolmark examiners has been remiss in searching for a statistical basis for evaluating their evidence is categorically untrue. Firearm and toolmark examiners, along with all other forensic scientists, have always recognized that behind every opinion there is an ultimate underlying statistical basis. To suggest otherwise would be intellectually dishonest. And it is beyond cavil that the field does not have a robust statistical model for firearm and toolmark evidence that is satisfactory in every respect. Statistical models have been created that provide some valuable insight into the nature of striated tool marks, but each model has its limitations, and it has not yet been established how relevant they will be to the universe of firearms and toolmark issues. But neither do we have a robust model for fingerprints, for trace evidence of various sorts, for bloodstain patterns, or for shoeprints and tire tracks. Nor is there a robust model for the appointment of judges, or the prediction of the weather, or predicting our personal longevity, or the future criminal behavior of a convict.

Substantial efforts have in fact been made within the community of firearm and toolmark examiners to establish a statistical basis for this type of evidence, to develop probability statements and error rate estimates. Many critical statements, like those illustrated above, have been made without acknowledging the significant headway that has been made on error rate estimates and measurement uncertainty by firearm and toolmark examiners in recent peer-reviewed literature. These have included studies of RMP, the statistical test mentioned by Carriquiry. Sadly, only anecdotal opinions and superficial reviews of select studies are offered up as informed criticism (5,6). Although these criticisms have been authored ostensibly by statisticians, no careful definitions of RMP as related to firearm and toolmark evidence have been proposed and no actual toolmark comparison research has been performed to support these criticisms. The purpose of the present article is in part to aid those who wish to understand the meaning of RMP in order to develop informed opinions on its utility for forensic firearm and toolmark identification. The present article is also intended as a brief review of the quantitative research into estimating RMP and error rates that has been published over the last several years.

In the United States, the "Daubert decision" has brought to the forefront the question of whether scientific theories, methodologies, or techniques have a "known or potential error rate." For a method that seeks to put an identity to a pattern of unknown origin, an error rate is most generally defined as how often a mistake can be expected to be made in the identification process. Such a statistic is not unique to firearm and toolmark examination, and there is currently much debate in the forensic science literature as to what "number" is the most important to compute (7–10). Straightforward system error rates can be estimated by computing the number of wrong identifications made on a sample of patterns drawn from a larger population of potential decisions. That process is anything but trivial. When former U.S. Supreme Court Chief Justice Rehnquist spoke of a "known or potential error rate" in his dissent to the Daubert decision, he may have been thinking of tallying up the number of mistakes and dividing that by the total number of cases where a conclusion was reached. He probably had no inkling as to the challenge that this would create on the part of those whose professional responsibility it would be to engage in such murky mathematics, or whether such calculations were in fact feasible.

A statistical test which has featured prominently in DNA profiling is RMP. Let us give consideration to this approach. In the context of firearm and toolmark identification, the random match probability would be the quantitative chance that two bullets, cartridge cases, or non-firearm-produced toolmarks will be identified as having been made by the same tool working surface given that they were in fact fired from different firearms, or made by different tools. In order to be able to estimate RMP and error rate statistics in general, data (objective or subjective) will be required. Toolmarks of all types are composed of microscopic "features." As described by Moran and Murdock (11) and reported in Neel and Wells (12), two-dimensional (2D) and three-dimensional (3D) toolmarks are characterized by features that:

- Occupy only the very surface of a recording medium (2D)
- Have been made in a very thin recording medium (2D)
- Result from the application of the tool to the medium in such a way that only superficial markings are produced (2D)
- Display discernible depth because the medium the toolmark is in has been displaced (3D)

From a statistical or data-driven point of view, the features that make up "evidence" correspond to random or near random variables that are distributed according to probability densities or mass functions. For the DNA profiling problem, the features are discrete alleles at well-defined loci that are (assumed to good approximation) statistically independent. Alleles have probability mass functions that can be easily estimated from population data. Unfortunately, a vanilla (frequentist) probabilistic-based analysis of features for toolmarks, analogous to DNA profiles, is not appropriate and out of the question. The continuous, nonbinary, and high-dimensional nature of toolmark surface features makes their probability density functions almost impossible to estimate accurately. This point has eluded many critics who view the application of RMP as the statistical prescription for the inherent uncertainty associated with firearm and toolmark evidence. RMPs cannot be estimated as easily or in the same way they can be for the DNA problem, precisely because of their spectrum of variability of the data. Drawing a direct comparison between the two problems is scientifically improper and not warranted. This is clear to researchers and practitioners in the forensic firearm/toolmark and DNA fields who have actually collected data to analyze.

However, it is incumbent upon the community of firearm and toolmark examiners to make an honest and continuing effort to estimate identification error rates or even RMPs. The need for this was underscored in 2009 by the National Academy of Sciences report on strengthening forensic science in the United States, which recommended "The development of quantifiable measures of uncertainty in the conclusions of forensic analyses" (13). This has been attempted in forensic firearm and toolmark identification, but it must be recognized that an error rate or an RMP is a mathematical function. They may be calculated given a statistical model and according to conventional rules of mathematics, but an assessment of their relevance is another matter. Camus (14) has cogently pointed out that "justice is not dispensed in a test tube, even a graduated one." In short, while the profession of firearm and toolmark examiners yearns for, and has striven for, a realistic probability model to explain firearm and toolmark individuality, a totally satisfactory and validated model has not yet been derived. But let us look at the progress that has been made to date.

Due to the structure of the problem, different analogies to other identification fields, outside of DNA profiling, must be sought. Saunders, Davis, and Buscaglia have clearly outlined an analogy with a common approach from biometric identification (15–17). False match error is an estimate of the (frequentist) probability that two toolmarks from different tools are declared to match. They go on to define RMP as "the chance of randomly selecting two (different tools) and then randomly selecting two (toolmarks), one from each (tool's) available (exemplar set), and declaring them to 'match' on the basis of the chosen comparison procedure... RMP is one measure of false match error" (15). While RMP, that is, false match error as defined in biometric identification, can be and has been estimated from available toolmark data, there is no clear indication that this approach offers general applicability to the universe of firearm and toolmark evidence. There are also existing studies that postulate theoretical models for toolmark RMP. These papers are published but have been ignored in recent affidavits attacking the veracity and scientific acumen of toolmark identification (3,18–25).

Review of the Current Literature in Statistical Applications for Firearms and Toolmarks

The firearm and toolmark identification community has not been indifferent toward the need to develop statistical estimates for toolmark individuality. As with other areas of forensic science, firearm and toolmark examiners have not lacked the wit to develop these statistics nor through laxness have failed to achieve totally suitable statistics. As early as 1932, Gunther and Gunther took the position that the needed probability factors must be established by research conducted by competent researchers (26). In 1935, both Hatcher, and Gunther and Gunther also described what data would be needed to develop RMPs for firearms and toolmarks (26–28). Hatcher went so far as to assign hypothetical numerical values to markings on bullets which resulted in a calculation for determining the chance of finding a firearm at random that would duplicate the markings on a questioned bullet. Hatcher went on to say:

> Of course, the details given in this discussion are all purely speculative, for no data are available as to the exact probability of the existence of any particular mark at any definite location on a bullet; but the discussion is intended to show how the mathematics of the theory of probabilities work in a simple case.

In 1949, O'Hara and Osterburg, after assuming that a statistical study had established the frequency with which certain toolmark characteristics occur, assigned numerical values to matching striated toolmarks (29). Churchman also reported that "tests have shown...," and then proceeded to give numerical values for the maximum frequency of occurrence for both striated and impressed toolmarks which he used to calculate probability estimates for the chance that a tool other than the one he identified could leave matching toolmarks. Unfortunately, Churchman provided no information, in his article or anywhere else, about how he developed his frequency of occurrence values (30).

In 1959, Biasotti recognized the need for empirically based statistical studies on toolmark patterns in a landmark study that has proven quite durable (31). His study recorded the number of groups of consecutive matching striations (CMS) between bullets fired from the same gun and bullets fired from different guns. Using these findings, he was able to estimate the probabilities of CMS run lengths of various sizes for known matching (KM) striation patterns and known nonmatching (KNM) striation patterns. Decades later, in 2007, Neel and Wells carried out a major extension of Biasotti's work where they made 4,188 comparisons of 2D and 3D striated toolmarks of assorted origin (12). Using tables of empirical counts for zero to greater than eight runs of CMS between two striae (2×) and greater than ten striae (10×), they estimated probabilities for each CMS run size. *Z*-tests were then used at the 99% level of confidence to compare the empirical frequencies for CMS run sizes between KM and KNM for both 2D and 3D striation patterns. Specifically, the best KNM CMS run sizes were compared to the most conservative KM CMS run sizes to determine whether there was evidence to reject the null hypothesis, that is, the proposition that the probabilities of these CMS run sizes are the same. In fact, evidence to reject the null hypothesis was detected as statistically significant differences were found. Neel and Wells also applied regression analysis, using observed KNM CMS run size probabilities to calculate hypothetical (future) KNM CMS run size probabilities. Using Bayes's theorem, they determined the likelihood ratios (LRs) of a KNM at different run sizes by dividing the estimated KM CMS run size probabilities by the KNM CMS run size probabilities (both actual and hypothetical). The LRs for KNMs for large CMS run sizes (10× or greater) were extremely large (meaning a KNM would be practically impossible), whereas the LRs for smaller CMS run sizes, 8× and lower, were relatively small. Neel and Wells do note that the LRs were determined with hypothetical KNM values and that a 2D CMS run larger than 6× and a 3D CMS run larger than 4× have never been observed or documented. Most helpfully, Neel and Wells published the CMS count tables obtained in their study. These tables can be used to educate new practitioners and be incorporated into other exploratory data analyses, such as fully Bayesian data analysis. For example, a classic

Bayesian analysis of this kind of count data is to use multinomial likelihoods with Dirichlet priors to ultimately yield RMP estimates.

In 2008, a paper by Howitt, Tulleners et al. (32) was published in which they proposed a model for the computation of "correspondence probabilities" (i.e., RMPs) between striation patterns imparted on bullets fired by the same gun versus bullets fired from different guns. For consistency of observations, the authors used the same functional definition of a "line" in a striation pattern as set forth by Biasotti. This definition makes their theory very general and equally applicable to striation patterns found in nonfirearm toolmarks as well as on firearms-related evidence. Their theory can also take into account arbitrary magnification levels (which would be a parameter of an optical comparison microscope) and the number of lines found in a striation pattern. An output of this approach is the probability of various CMS runs on striation patterns generated by different sources matching purely by chance. The Howitt–Tulleners study computed that the probability of random correspondence between 2× CMS runs on bullets (called "doublets" in the study) from different sources is between 0.1–0.16 at 20 μm resolution (i.e., at 10–16% chance) and 0.14–0.24 at 30 μm resolution. Biasotti's empirically derived probabilities for the same situation are 0.2–0.46, depending on whether or not the bullet is copper-jacketed. The theoretical RMPs for 3× runs are 0.003–0.005 at 20 μm resolution and 0.007–0.01 at 30 μm resolution. Biasotti's empirically derived probabilities closely agree at 0.01–0.1, depending on jacketing (31,32).

There have been two notable attempts to develop alternative RMP/identification error rate estimates for impression toolmarks. The first, by Stone, was theoretical, and the second, by Collins, was an empirical validation of Stone's work (33,34). Both studies demonstrated that valid impression toolmark identifications could be made using a relatively small number of individual toolmark features, but neither study resulted in examiners being able to calculate case-specific RMPs.

With the advent of confocal microscopy and laser scanners, the acquisition of the entire 3D surface of a toolmark can now be obtained. A study by Bachrach, Jain, Jung, and Koons documented the use of confocal microscopy to digitally record the surfaces of striated toolmarks made by screwdrivers and tongue and groove pliers (35). Similarity scores for all possible comparison pairs of signatures were generated based on the cross-correlation function and used to produce matching and nonmatching score distributions (histograms). Algorithm-generated identifications were found to be highly reliable as long as the screwdrivers' angles of tilt relative to the substrate surface were consistent (angle of tilt was obviously not an issue for tongue and groove pliers). In these cases, at a 45° angle of tilt, estimated RMPs (denoted as "empirical error rates" in the paper) were 0.0011 for screwdriver striation patterns and 0.0003 for tongue and grove pliers.

Chumbley et al. authored a series of papers comparing the effectiveness of their comparison computer programs to human examiners (36–38). These programs use 3D toolmark surface data from focus variation microscopy combined with an identification algorithm to identify a region of best agreement between two toolmark datasets being compared. The algorithm searches for regions of best fit (on both toolmarks) and compares correlation values. If a match exists at one point along the scan length, there should be large correlations along their entire length of the toolmark. RMPs (called "error rates" in the paper) estimated using this method were 0.023 for toolmarks made at a 30° angle of tilt and 0.01 at 60° and 85° angle of tilt (37). The authors then conducted a double-blind study in which fifty experienced toolmark examiners gave their opinions on the sample set. In summary, the authors determined that examiner performance was much better (lower error rate) than the algorithm, but the identified deficiencies in their automated method are now susceptible to being addressed and improved upon.

Chu et al. (39) describe a procedure for automated bullet signature identification using confocal microscopy and correlation calculations. This procedure has the ability to automatically select the effective correlation area, calculate the twist angle, extract an average profile, and filter out information that does not resolve individual characteristics, with the intent to produce higher correlation ratios for matching bullets.

In a separate article, Chu et al. (40) introduce striation density (d_s) as an objective criterion for quantifying the suitability of bullet images for automatic bullet signature correlation. Their results indicated that there was a strong relationship between striation density and identification rate. Striation density was found to be strongly affected by the image quality. Expanding on this study, Chu et al. (41) introduced a method of valid correlation area selection based on striation edge detection. They found that the threshold length (L) has a significant effect on distinguishing a valid striation area from an invalid one, as well as the calculation of the density striation parameter (d_s). Compared with their previous study (40), the group found an 8.8% improvement in the classification results.

In yet another study, Chu et al. (42) employed a model using quantitative consecutive matching striae identification criteria (QCMS) in the 3D confocal comparison of bullets fired from consecutively rifled barrels where known standards and blind samples were intercompared. The system automated the selection and masking of nonstriae surface features, and informative striae were compared using both a mathematical cross-correlation statistical algorithm and a QCMS profile system that counted consecutive striae. Both approaches accurately separated known matching and known nonmatching bullets.

More recently, Weller et al. (43) used confocal microscopy to analyze breech face markings on 90 cartridge cases test fired from ten consecutively manufactured pistol slides. The data from this study, when plotted in a histogram, appear to be normal distributions. There is clear separation between the matching and nonmatching test fires. If the data are theorized to be from a larger dataset with exponentially thin tails, there is a small overlap between them because the probabilities represented by the tails never actually reach zero. Subsequent to the publication of his article, Weller used his data to calculate a RMP. In his model, Weller set a threshold at the lowest scoring known match cross-correlation score (CCF) of 0.57. Based on the distributions of matching and nonmatching confocal data, Weller picked this point because it was well above the highest scoring known nonmatch CCF and was at the lower end of the known matching data (thus conforming to the AFTE Theory of Identification). Then, assuming a normally distributed nonmatching dataset, the probability of a nonmatch comparison reaching the chosen threshold was calculated to be 3.00×10^{-35} (or 1 in 3.34×10^{33}). The threshold (0.57 CCF) that Weller chose for his RMP calculation was also based on the point which he felt a firearm examiner could mistake a nonmatch for a match because the surface geometry would be similar to other known matches and greater than the best known nonmatch due to the small theoretical overlap in the tails of the two distributions. Based on Weller's data, although there is a statistical possibility of such an overlap, the practical chance of this occurring is extremely small, as evidenced by the calculated RMP. Weller presented this RMP calculation at the May 2011 California Association of Criminalists meeting in Long Beach, CA. He stressed that this RMP only applied to his data and samples and should not be expanded to other firearms where different manufacturing processes may have been used. As mentioned above, this calculation assumes normally distributed nonmatch data: a distribution curve with exponentially thin tails. Other distributions are likely better models for the data, and this is currently being researched (unpublished data). However, the calculated probability was presented not to provide a unifying model for calculating random match probabilities, but only to convey the substantial separation of nonmatch and matching distributions.

In a 2012 study, Petraco et al. (44) applied multiple statistical pattern recognition methods to 75 striation marks made by nine identical high-quality Craftsman brand quarter-inch slotted screwdrivers. Striation patterns were represented by barcodes and used to fit robust identification models known from the field of machine learning. Fits for these models showed the conservative upper limit of the RMP (called "error rate" in the paper) was well below 0.02. The authors also gave instructions on how to assign rigorous levels of confidence to specific toolmark identifications.

Gambino et al. (45) applied the same machine learning methods to 58 primer shear marks on 9 mm cartridge cases fired from four Glock model 19 pistols. The striation patterns were scanned with a confocal microscope and preprocessed with metrological software. The results were very promising, showing very low error rates. The sample size was then significantly increased in a follow-up study and also included screwdriver striation patterns (46). A database was assembled consisting of 290 3D screwdriver striation patterns generated by 29 screwdrivers and 162 3D primer shears generated by 24 Glock pistol slides. A stochastic simulator for striation pattern signatures was also developed to increase the database to arbitrary size and identification difficulty. Using the real data along with the simulator results, RMP estimates were 0.0001 for screwdrivers (1740 striation patterns used) and 0.0003 for Glock primer shears marks (720 patterns used). The authors have also made all of their raw data and analysis programs available for scientists and practitioners to use or expand upon in their own research.

Hamby, Norris, and Petraco performed an analysis of a set of 1632 9 mm Glock fired cartridge cases collected over a 21-year period (47). The experiment subjected the set to manual comparisons of the firing pin aperture shear on the primers using traditional pattern matching. In total, 1,330,896 pairwise comparisons were made. Each firing pin aperture shear was found to be distinct from every other and uniquely identifiable. A Bayesian methodology was then applied to estimate an upper limit for a random match probability when comparing these types of surfaces. The RMP estimate was 0.0001% (47). A subset of 617 cartridge cases was also subject to comparison by the IBIS correlation algorithm (190,036 pairwise comparisons). Again, no cartridge case was misidentified for any other and a similarly vanishingly small RMP was estimated, consistent with the manual comparison estimate.

Current State of the Use of RMPs in Firearm and Toolmark Identification

To summarize the statistical issues—at the present time, with the exception of quantitative consecutive matching striae (QCMS) identification criteria values for striated toolmarks—there is no unifying model for the comparison of striated and impressed toolmarks as there is in the field of forensic DNA profiling. Recent computation studies have in fact attempted to estimate RMP (usually simply referred to as "error rates" in those papers). While those estimates do vary, one fact is clear: Whatever the RMP is, the error rate is very low.

It should also be recognized that whatever the "known or potential error rate" is for firearm and toolmark identification, it is extremely low when compared to other error within the judicial system. For an example, we are informed that upon appeal, typically 20–30% of cases are reversed or remanded because of judicial error. It may be unproductive to compare error in evidence evaluation and interpretation with other types of error, but clearly we are operating in a milieu in which error is anything but absent. The community of firearm and toolmark examiners recognizes a professional obligation to measure error, and to manage it. It is not clear that all other nonscientific elements of the justice system are as attentive to this level of accountability.

Furthermore, it is not known at this time whether it is altogether feasible to apply RMP to the identification of firearms and toolmarks in a manner that will satisfy everyone. As D. Michael Risinger, a highly respected academic observer and critic of the comparative forensic sciences since the early 1980s commented (48):

> There is one school of thought in the academy (inferred to be the National Academy of Science) that holds, essentially, that without quantified random match probability data, all expert claims concerning identification or attribution should be regarded as invalid. But I believe this goes too far. Random match probability modeling for phenomena like toolmarks or handwriting is many times more complicated than for DNA, where we are blessed with an underlying system that comes in basically binary units, and DNA has been hard enough to get control of.

Rudin and Inman observed that constructing useful and relevant databases for nonbiological evidence is appreciably more challenging than for biological evidence (49). They also commented that since fundamental differences exist in the nature of nonbiological evidence and the dynamics of source populations, it is impossible to directly apply the DNA typing model.

RMPs can be used in DNA profiling because analysts base their estimates of association on arrays of genotypes, which are analogous to subclass characteristics in firearm and toolmark identification (50). Firearm and toolmark identifications are based on arrays of individual, not subclass, characteristics. Another reason why RMPs are so difficult to establish for firearm and toolmark identification is that tool working surfaces, many of which are microscopically unique when they are produced, may not remain stable over time; the working surfaces may continually and unpredictably change, through use and wear in consideration of the career and history of the tool or the firearm. Additionally, the microscopic marks used by firearm and toolmark examiners on a particular tool may originate from a variety of machining processes. For this reason, it will be extremely difficult to develop a universal mathematical model that can accurately predict the random toolmarks left by some tools. For example, if a tool's working surface is sanded with sandpaper, the striated marks left can depend on the grit of sandpaper, the amount of pressure used by the sander, the wear of the sandpaper, and the hardness of the tool working surface. Minor changes to these factors will affect the depth, size, density, and frequency of the striated marks left behind. When one contrasts this example with the mostly binary inheritance ($p^2 + 2pq + q^2$) patterns of DNA, the authors question if the critics of firearm and toolmark examination understand and fully appreciate the task at hand. In firearm and toolmark examination, often the exception is the rule. Therefore, the authors approach probability models for firearm and toolmark evidence with due scientific caution and skepticism.

Despite the difficulty in presenting firearm and toolmark examination in a probabilistic manner, the research cited in this article has great use to the profession. While none of these examples provide a method for presenting probability in individual casework, they do provide context for nonfirearm examiners when evaluating “identifications.” When the AFTE Theory of Identification was adopted, the firearm and toolmark examiner profession described the chance of a random match as a “practical impossibility.” The research reviewed in this article provides strong mathematical and objective support for that hypothesis and conclusion.

Kaye et al. (51) fairly describes the forensic firearm and toolmark examiner’s expertise as the training and “experience-based ability to assess, analyze,” and compare toolmarks “to differentiate signal from noise” (individual toolmarks from spurious detail) “and artifact from discrepancy” (insignificant from significant differences), “and to evaluate whether two patterns” (toolmarks—questioned compared to test) “did or did not come from the same source”. This same source determination is made to the practical, and not absolute, exclusion of other tools. This distinction between practical and absolute identification was described as far back as 1929 by Derome when he made the observation that “the certainty which this evidence (firearm identification) bring with it does not extend to the absolute, since the absolute exists only in mathematics; it is a physical certainty like that which our senses give us. . .”[1]

Absolute Versus Practical Identification and Subjectivity

Derome’s comment, made 85 years ago (as of this writing), enables us to segue to another criticism often levied at firearm and toolmark identification: the issue of absolute identification and practical certainty. Critics have often misrepresented the position of firearm and toolmark examiners by declaring that examiners claim absolute identification, with statements such as “the evidence bullet was fired by the suspect’s firearm to the total exclusion of all other firearms in the world.” This is a straw man, and should be recognized as such. Absolute certainty opinions may have been adopted in the past, but this type of position has been retired for some time and no longer represents the consensus thinking of the firearm and toolmark community. Practical certainty has a place in the resolution of conflict just as it has a place in our everyday lives. And our everyday lives are predicated upon practical certainty. There is a practical certainty that our car will start in the morning (assuming it is in good mechanical condition), or that our (normally obedient) dog will come when called. Practical certainty should not be allowed to be seen as an enemy of justice. Firearm and toolmark examiners have struggled with how to accurately express the certainty of opinions as related to the uniqueness of the evidence with which they are confronted. But then others have struggled as well with the concept of uniqueness—philosophers, Boolean algebra mathematicians, and rare stamp dealers. Ultimately, we all must make peace with this elusive concept. In addressing this criticism, one is left to wonder why it is that the legal academics who have been vocal on this issue seem to have no problem with the “reasonable doubt” concept, yet are mystified and/or outraged at the collateral “reasonable (practical) certainty” concept.

Nor should the subjective aspects of firearm and toolmark identification be seen as a grave or mortal flaw in firearm and toolmark examination. The subjectivity that is currently unavoidable in firearm and toolmark identification comes at the end of a number of quantifiable and measurable steps in the examination process. For instance, bullet weights, caliber measurements, land and groove counts, land width measurements, and direction of twist are all nonsubjective observations and measurements that greatly narrow down the universe of potential sources for a fired bullet. Likewise, the use of an optical comparison microscope to compare surface features of two toolmarks is an objective method, where known and questioned areas are visually compared, but not measured, directly. Those areas may be photographed or examined by others. The subjective process begins when the examiner judges if any similarity is sufficient for identification, or not. Critics often treat “subjectivity” as a pejorative word, suggesting that an opinion derived from a subjective evaluation is bereft of validity. The issue of subjectivity is mentioned in virtually every critical review of firearm and toolmark evidence, and to an ever increasing extent is creeping into court decisions as well (52).

We must remind ourselves that we may be diagnosed with influenza by a doctor after 60 seconds of a very subjective examination based on the doctor’s training and experience, or that we pay a good deal of money to an automobile mechanic who has made a very subjective determination of some mechanical problem, or that we have identified our own toothbrush in the morning based on totally subjective features, or that we identify our spouse in a crowded airport based largely on subjective features. To use a more germane example, the identification of an infrared spectrum of a drug or explosive recovered from some form of physical evidence is also largely subjective. The best infrared (IR) spectrum ever obtained from an evidence sample will never exactly “match” the standard spectrum. Indeed, if an analyst were to ever see two IR spectra like this that did, they should strongly suspect one has been mislabeled “evidence” and the other “reference standard” when, in fact, they are two copies of the same spectrum. Hence, there are subjective judgments in the field of analytical chemistry and compound identification with IR spectroscopy (as well as other techniques, such as GC-MS), as there are with many other areas of science. As with practical certainty, actions and opinions derived from subjective observations have a place in a system of justice just as they do in our everyday lives.

References

1. Daubert V. Merrell Dow Pharmaceuticals, Inc. - 509 U.S. 579 (1993).
2. Schwartz A. Challenging firearms and toolmark identification - part 1. Champion 2008;44:10–8.
3. Carriquiry A. Declaration of Alicia Carriquiry in support of Joseph Blacknell’s motion to exclude firearms and toolmark identification evidence, or in the alternative, for a Kelly hearing, 2012; https://afte.org/

---

[1] Derome W. Expertise en armes à feu (A treatise on firearms). Montreal, 1929. Unpublished manuscript of French-to-English translation by Alfred E. Riedel, Berkeley (California) Police Department; July 24, 1931. Page 59 (corresponds to page 69 of original text). A copy of the translated manuscript is available upon request.

uploads/documents/swggun-cavblacknell-carriquiry.pdf (accessed June 25, 2016).

4. CA v. Miller; Court ruling and order to exclude the testimony of Alicia Carriquiry. Superior Court of the State of California, Case No. 62-98243 (2012).
5. Spiegelman C, Tobin WA. Analysis of experiments in forensic firearms/toolmarks practice offered as support for low rates of practice error and claims of inferential certainty. Law Proba Risk 2013;12:115–33.
6. Tobin WA, Blau PJ. Hypothesis testing of critical underlying premise of discernible uniqueness in firearms-toolmarks forensic practice. Jurimetrics 2013;53:121–46.
7. Petraco N. Color atlas of forensic toolmark identification. Boca Raton, FL: CRC Press, 2010.
8. Taroni F, Aitken CCG, Garbolino P, Biedermann A. Bayesian networks and probabilistic inference in forensic science. Hoboken, NJ: Wiley, 2006.
9. Koller N, Nissen K, Ries M, Sadof E. Probability conclusions in expert opinions on handwriting. substantiation and standardization of probability statements in expert opinions. Munich, Germany: Luchterhand, 2004.
10. Taroni F. Inadequacies of posterior probabilities for the assessment of scientific evidence. Law Proba Risk 2005;4:89–114.
11. Moran B, Murdock J. The application of the scientific method to firearm and toolmark examination. AFTE J 2003;35(2):234–40.
12. Neel M, Wells M. Comprehensive statistical analysis of striated tool mark examinations part 1: comparing known matches to known non-matches. AFTE J 2007;39(3):176–98.
13. National Research Council. Strengthening forensic science in the United States: a path forward. Washington, DC: National Academies Press, 2009;23.
14. Camus A. Reflections on the guillotine. Michigan City, IN: Fridtjof-Karla Publications, 1959.
15. Saunders CP, Davis LJ, Buscaglia J. Using automated comparisons to quantify handwriting individuality. J Forensic Sci 2011;56(3):683–9.
16. Wayman JL. Confidence interval and test size estimation for biometric data. San Jose, CA: Department of Defense, 2000; Report No: MDA904-97-C-03.
17. Schuckers M. Computational methods in biometric authentication. New York, NY: Springer, 2010.
18. Tobin W.A. Declaration of William Tobin. US v. Alls 2009; https://afte.org/uploads/documents/swggun-usvalls-tobin.pdf (accessed June 25, 2016).
19. Tobin W. Affidavit of William A Tobin. State of Arizona v. William Wayne Macumber 2011; https://afte.org/uploads/documents/swggun-azvmacumber-tobin.pdf (accessed June 25, 2016).
20. Nixon J. Affidavit of John Nixon. State of Arizona v. William Wayne Macumber 2011; https://afte.org/uploads/documents/swggun-azvmacumber-nixon.pdf (accessed June 25, 2016).
21. Saks M. Affidavit of Professor Michael J Saks. State of Arizona v. William Wayne Macumber 2011; https://afte.org/uploads/documents/swggun-azvmacumber-saks.pdf (accessed June 25, 2016).
22. LaMagna D. Preliminary Affidavit of David J LaMagna in support of defendant's Daubert motion. State of Texas v. Nguyen, Thai-An Huu 2011; https://afte.org/uploads/documents/swggun-txvnguyen-lamagna.pdf (accessed June 25, 2016).
23. Spiegelman C. Declaration of Clifford Spiegelman in support of Joseph Blacknell's motion to exclude firearms and toolmark identification evidence, or in the alternative, for a Kelly hearing 2012; https://afte.org/uploads/documents/swggun-cavblacknell-spiegelman2.pdf (accessed June 25, 2016).
24. Tobin W. Testimony of William Tobin. US v. Clifford D Jackson and Clifford Durham Jr. 2012; https://afte.org/uploads/documents/swggun-usvjackson-durham-tobin.pdf (accessed June 25, 2016).
25. Tobin W. Testimony of William Tobin. State of Florida v. Shawn Emmanuel Richardson, 2013; https://afte.org/uploads/documents/swggun-flvrichardson-tobin.pdf (accessed June 25, 2016).
26. Gunther C. Markings on bullets and shells fired from small arms. Mec Eng 1932;54:341–5.
27. Hatcher JS. Textbook of firearms investigation, identification, and evidence. Plantersville, SC: Small-Arms Technical Publishing Company, 1935;274–89.
28. Gunther JD, Gunther CO. The identification of firearms: from ammunition fired therein with an analysis of legal authorities. New York, NY: John Wiley & Sons Inc, 1935.
29. O'Hara CE, Osterburg JW. An introduction to criminalistics: the application of the physical sciences to the detection of crime. New York, NY: Macmillan Co., 1949;666–79.
30. Churchman JA. Engraving & impressions. RCMPQ 1949;14(3):136–41.
31. Biasotti A. A statistical study of the individual characteristics of fired bullets. J Forensic Sci 1959;4(1):34–50.
32. Howitt D, Tulleners F, Cebra K, Chen S. A calculation of the theoretical significance of matched bullets. J Forensic Sci 2008;53(4):868–75.
33. Stone R. How unique are impressed toolmarks? AFTE J 2003;35(4):376–83.
34. Collins E. How unique are impressed toolmarks?: an empirical study of 20 worn hammer faces. AFTE J 2005;37(4):252–95.
35. Bachrach B, Jain A, Jung S, Koons RD. A statistical validation of the individuality and repeatability of striated toolmarks: screwdrivers and tongue and groove pliers. J Forensic Sci 2010;55(2):348–57.
36. Faden D, Kidd J, Craft J, Chumbley L, Morris M, Genalo L, et al. Statistical confirmation of empirical observations concerning toolmark striae. AFTE J 2007;39(3):205–14.
37. Chumbley L, Morris M, Kreiser M, Fisher C, Craft J, Genalo L, et al. Validation of tool mark comparisons obtained using a quantitative, comparative, statistical algorithm. J Forensic Sci 2010;55(4):953–61.
38. Chumbley L. Quantification of toolmarks, final technical report. Washington, DC: NIJ, 2010; Report Number: 2004-R-IJ-088.
39. Chu W, Song J, Vorburger T, Yen J, Ballou S, Bachrach B. Pilot study of automated bullet signature identification based on topography measurements and correlations. J Forensic Sci 2010;55(2):341–7.
40. Chu W, Song J, Vorburger T, Ballou S. Striation density for predicting the identifiability of fired bullets. J Forensic Sci 2010;55(5):1222–6.
41. Chu W, Song J, Vorburger TV, Thompson R, Silver R. Selecting valid correlation areas for automated bullet identification system based on striation detection. J Res Natl Inst Stand Technol 2011;116(3):647–53.
42. Chu W, Thompson R, Song J, Vorburger T. Automatic identification of bullet signatures based on consecutive matching striae (CMS) criteria. Forensic Sci Int 2013;231:137–41.
43. Weller T, Zheng A, Thompson R, Tulleners F. Confocal microscopy analysis of breech face marks on fired cartridge cases from 10 consecutively manufactured pistol slides. J Forensic Sci 2012;57:912–7.
44. Petraco NDK, Shenkin S, Speir J, Diaczuk P, Pizzola P, Petraco N. Addressing the National Academy of Sciences challenge for toolmarks: a method for statistical pattern comparison of striated toolmarks. J Forensic Sci 2012;57(4):900–15.
45. Gambino C, McLaughlin P, Kuo L, Kammerman F, Shenkin P, Diaczuk P, et al. Forensic surface metrology: tool mark evidence. Scanning 2011;33(5):272–8.
46. Petraco NDK, Kuo L, Chan H, Phelps E, Gambino C, McLaughlin P, et al. Estimates of striation pattern identification error rates by algorithmic methods. AFTE J 2013;45(3):235–44.
47. Hamby J, Norris S, Petraco NDK. Evaluation of GLOCK 9 mm firing pin aperture shear mark individuality based on 1,632 different pistols by traditional pattern matching and IBIS pattern recognition. J Forensic Sci 2016;61(1):170–6.
48. Risinger D. The NAS/NRC report on forensic science: a path forward fraught with pitfalls. Utah L Rev 2010;2:243–4.
49. Rudin N, Inman K. The shifty paradigm, part 1: who gets to define the practice of forensic science?The CAC News 2006;4th Quarter:13–6; http://www.forensicdna.com/assets/4thq05.pdf
50. Gutowski S. A response to: a systemic challenge to the reliability and admissibility of firearms and toolmark identification, a recently published article by Adiana [sic] Schwartz. The Forensic Bulletin (ISSN 1447–6673; National Institute of Forensic Science Australia), Winter 2005;22–6; http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.116.7654&rep=rep1&type=pdf
51. Kaye D, Bernstein D, Mnookin J. Qualitative testimony on forensic science identification (footnote 21). In: The new Wigmore: a treatise on evidence: expert evidence, 2nd edn. New York, NY: Aspen Publishers, 2011;689.
52. U.S. v. Glynn, 578 F. Supp. 2d, 567 (2008).

Additional information and reprint requests:
John E. Murdock, M.S.
2530 Arnold Drive, Suite 200
Martinez, CA 94553
E-mail: jmurdock@so.cccounty.us