Received: 00 Month 0000 | Revised: 00 Month 0000 | Accepted: 00 Month 0000

DOI: xxx/xxxx

Statistical Methods for Assessing Strength of Cartridge Case Evidence

Fredrick Osei

[1]Department of Statistics, Iowa State University, Iowa, USA

Abstract

The quantification of evidence in forensics is crucial for investigating crime and its consequent verdict by a jury. Two cartridge cases can be classified as a match (fired from the same firearm) or non-match (not fired from the same firearm) using the number of similar regions obtained when they are compared. The Congruent Matching Cells (CMC) algorithm generates a score representing the number of valid correlation regions when comparing cartridge cases from different crime scenes. The CMC score only enables the classification of whether the compared cartridge cases match or do not match using certain proposed thresholds. That leads to a need for more necessary details as to the extent of match or non-match when comparing the defense and prosecution propositions at trial by a jury. We illustrate the application of statistical methods such as a two-stage approach, score-based likelihood ratio (SLR), and Bayesian approach (posterior model probability and Bayes factor) to assess the strength of evidence that cartridge cases from different crime scenes were fired from the same firearm. These statistical methods are an improvement in using only the CMC score to classify two cartridge cases as a match or a non-match. The performance of the statistical methods revealed at least a 98% average value of sensitivity, specificity, and accuracy based on out-of-sample cross-validation estimates using a cartridge cases scan database.

KEYWORDS:
Congruent Matching Cells (CMC); valid correlation region; cartridge case; common source; crime scene.

1 | INTRODUCTION

Statistical methods are used in forensics to assess the strength of evidence that materials such as cartridge cases provide. The application of statistical methods aids the quantification and analysis of evidence obtained from crime scenes during trial. Forensic experts collect cartridge cases at crime scenes, which are used to generate data known as the Congruent Matching Cells (CMC) scores using the CMC procedure. These scores are then analyzed to assess the strength of evidence for specific propositions. In court, the propositions involved are the prosecution and defense propositions. The jury evaluates the evidence presented during the trial to determine whether the evidence has more weight toward one of these two propositions. Thus, evaluating evidence enables the jury to decide on the strength of evidence towards the defendant's guilt. The jury's verdict stems from making a decision after an analysis of the evidence by an expert (Devine, 2012).

Markings on a cartridge case provide valuable forensic information (Song 2013). The process through which a breech face impression (markings) is left on a cartridge case begins when a cartridge is loaded into a firearm. The cartridge typically consists of a casing, a primer, a powder, and a projectile. When the trigger is pulled, the firing pin hits the primer, which ignites the powder inside the cartridge. This also forces the projectile out of the casing and down the barrel. The cartridge case is ejected from the firearm after firing. The cartridge case's marks are transferred from the breech face of the barrel through the firing (see Figure 1). Forensic experts collect the fired cartridge cases from the crime scenes for further analysis. According to Geradts et al. (2001), the breech face impressions on the recovered cartridge cases can be compared to known impressions from different firearms. If a match is found between the breech face impression on a recovered cartridge case and the known breech face impression on another cartridge case, it can link the firearm to a crime.

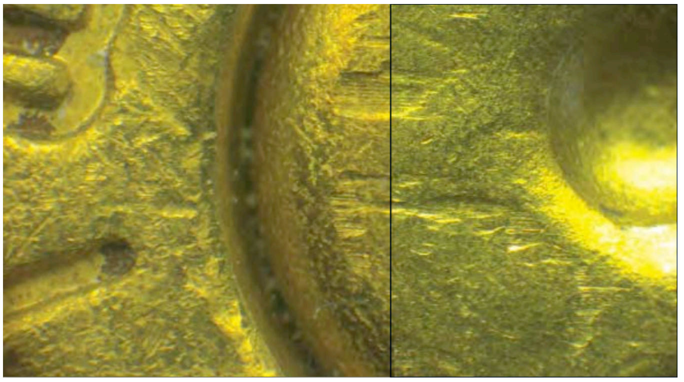

FIGURE 1 A cartridge case pair with visible breech face impressions under a comparison microscope. A thin line can be seen separating the two views. The extent to which the markings coincide is used to conclude that the pair match or do not match[4].

For a given pair of cartridge cases, the Congruent Matching Cells (CMC) algorithm is used to generate the CMC score between the two cartridge cases, which can then be used to determine whether the cartridge cases are a *match*—fired from the same firearm or *non-match*—not fired from the same firearm (Tong et al. 2014). That involves partitioning the reference cartridge case scan into a grid of cells. Each cell is compared to its maximum matching counterpart on the target cartridge case scan (see Figure 2). Partition of the reference cartridge case scan is done on the whole scan. However, only its valid correlation regions tend to be congruent on particular regions of the target cartridge case scan. Song (2013) defines *valid correlation regions* as regions where "the individual characteristics of the ballistic signature are found that can be used effectively for ballistic identification." Thus, the CMC procedure is only applied to valid correlation regions of cartridge case scans as denoted by the gray region in Figure 2 .

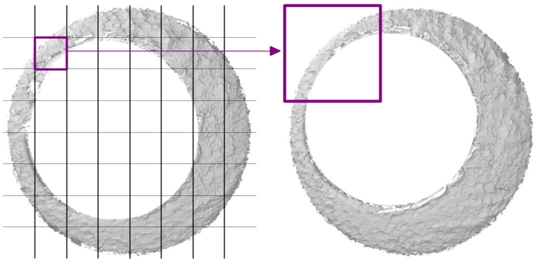

FIGURE 2 Partition of the reference cartridge case scan into 8 × 8 grid, leading to 64 cells. For each cell on the reference cartridge case, the target cartridge case is rotated to find its maximum matching counterpart. The number of matching cells is used as the output of the Congruent Matching Cells algorithm[4].

Considering that a jury wants to assess the strength of evidence that two cartridge cases from two crime scenes are a match, a particular statistician needs to analyze the data (CMC scores) obtained from the evidence materials (cartridge cases), and that will help the jury decide. This study adopted three statistical methods, including a two-stage approach, score-based likelihood ratios (SLRs), and Bayesian approach (posterior model probabilities and Bayes factor), which were used to analyze the CMC scores, thus enabling the jury to assess the strength of evidence of whether the two cartridge cases match. Specifically, what is the strength of evidence that the cartridge cases from the first crime scene and those from the second crime scene share a *common source* (fired from the same gun)? This research provides an approach to tackle such question. The common source objective was used instead of the *specifc source* (whether the cartridge cases were fired from the suspect's gun) because of the nature of the "example data" and few cartridge cases from each barrel in the database. We also compared the usage of these statistical methods for determining the strength of evidence obtained from breech face impressions on cartridge cases.

2 | DATA

This section describes the data generated by comparing the cartridge cases from the crime scenes and the processed cartridge case scans database.

We have two datasets: comparing the cartridge case from the first crime scene to the two cartridge cases from the second crime scene and the cartridge case scans database. Comparison of each cartridge case from the second crime scene to the cartridge case from the first crime scene using the CMC algorithm generates a CMC score. A cartridge case from the first crime scene and two cartridge cases from the second crime scene comprise two CMC scores in the first dataset. On the other hand, the cartridge case scans database contains CMC scores from a pairwise comparison of forty cartridge cases from ten barrels.

2.1 | Crime Scene Cartridge Cases Data

Table 1 shows the CMC scores comparing the cartridge case from the two crime scenes. The first crime scene cartridge case is the reference, and the cartridge cases from the second crime scene are the target. Description of the cartridge cases as reference and target are parameters for the CMC algorithm and the resulting CMC score is symmetric—we would get the same score if we switched which cartridge case was the "reference" and which was the "target". Each of the comparisons shown in Table 1 has a unique CMC score. However, it is possible for two different pairs of cartridge cases to have the same CMC score. The crime scene cartridge cases are tagged with CS and a number. For instance, CS-1 means it is a cartridge case from the first crime scene.

TABLE 1 Example Congruent Matching Cells (CMC) score (out of a maximum of 64) when comparing one crime scene cartridge case to two cartridge cases from a different crime scene. The CMC column shows the score when the first crime scene cartridge case is used as the target and the second crime scene cartridge case is used as the reference.

First Crime Scene Cartridge Case	Second Crime Scene Cartridge Case	CMC
CS-1	CS-2a	3
CS-1	CS-2b	7

2.2 | Processed Cartridge Cases Scans Database

The database is a data of matches and non-matches resulting from a pairwise comparison of forty cartridge cases. The database can be found from https://iastate.figshare.com/articles/dataset/Zemmels_et_al_2022_Cartridge_Case_Scans/19686297/1

The Reference and Target columns in Table 2 are descriptions of the cartridge cases that were compared. The first number denotes the barrel and the second number denotes the cartridge case label. For instance, 1-1 denotes first cartridge case from barrel 1, 1-2 denotes second cartridge case from barrel 1. Reference and Target cases with equal first numbers (1-1, 1-2) denote matches, and those with unequal first numbers (1-1, 2-1) represent non-matches.

The database includes 780 observations (from a pairwise comparison of 40 cartridge cases), 10 barrels, 63 matches, and 717 non-matches. A pairwise comparison of cartridge cases from the same gun is a match. So, the 63 matches are CMC scores from comparing cartridge cases from the same firearm. The 717 non-matches entail a pairwise comparison of cartridge cases from different firearms. The threshold (which is 4) for matches and non-matches in the database according to Figure 3 separates the two distributions. The threshold in Figure 3 is 4 because it separates known matches and known non-matches. Song (2013) proposes an actual threshold of 6.

TABLE 2 Sample cases of the processed database showing the Congruent Matching Cells (CMC) scores generated from a pairwise comparison of 40 cartridge cases from 10 barrels. The reference cartridge cases were partitioned into 64 cells and compared to regions of the target cartridge cases to find the number of matching counterparts.

Reference	Target	CMC
1-1	1-2	16
1-1	10-1	1
1-1	10-2	0
1-1	2-1	0
1-1	2-2	0
1-1	3-1	0
1-1	3-2	0
1-1	4-1	2

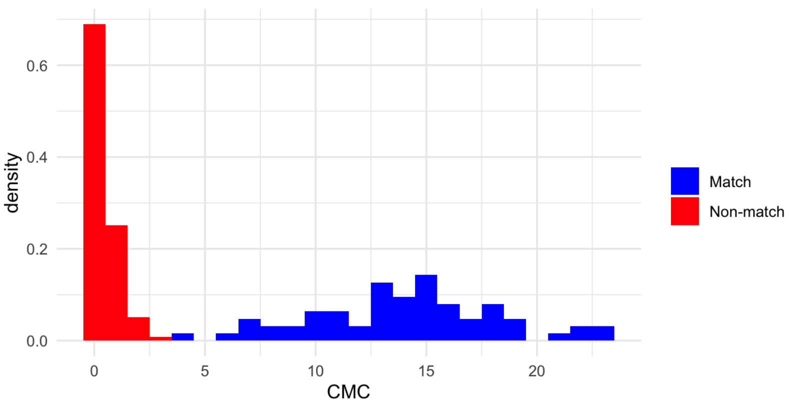

FIGURE 3 Histogram of the Congruent Matching Cells (CMC) scores in the processed cartridge cases database depicting two distributions: Matches and Non-matches. The number of scores in the non-matches' distribution exceed that of the matches. The total possible CMC score is 64, and the range for matches is 5 upward.

Figure 3 provides an overview of the ranges and distributions of CMC scores for matches and non-matches. The CMC scores for non-matches range from 0 to 4, and that of matches range from 5 to 26. Thus, the CMC scores for matches have more spread. A pair of cartridge case that are a match share a common source. There is no visibility of an overlap between the two distributions. Since there is no overlap, we have two likelihoods with independent parameters.

3 | METHODS

This section describes the chosen propositions, notations used for the datasets, and probability models. Also, the various methods employed for assessing the strength of evidence that cartridge cases from the crime scene were fired from the same gun are defined. Additionally, the performance metrics used to compare the methods are explained.

3.1 | Proposition

Suppose there were two crime scenes in Ames; a cartridge case (first unknown source cartridge case) was recovered from the first crime scene, and two cartridge cases (second unknown source cartridge cases) were recovered from the second crime scene. Scans of these cartridge cases from both crime scenes were conducted and passed through the CMC algorithm to generate the CMC scores to determine whether they were fired from the same firearm (whether there were one or two firearms involved).

The jury considers two propositions to make a decision. Using the common source forensic narrative as illustrated in Ommen and Saunders (2021), these propositions include:
Prosecution (H_p): The two sets of cartridge cases (unknown source evidence) originate from the same source (firearm).
Defense (H_d): The two sets of cartridge cases (unknown source evidence) originate from different sources (firearms)

3.2 | Data Notation

We used the following notations for the datasets:

1. Y^s is the CMC score for comparison of the cartridge case from the first crime scene and the cartridge cases from the second crime scene.
2. Y^D is the CMC scores for all cartridge case pairs in the database.
3. $Y^D = \{Y^M, Y^N\}$ denotes two populations in the database.
4. Y^M is the CMC scores for all matches in the database.
5. Y^N is the CMC scores for all non-matches in the database.

Probability models for CMC scores are constructed using the database. We utilized the distribution of matches and non-matches in the database. Although we have uneven number of matches and non-matches in the database, their respective amount is still adequate to generate models for the CMC scores.

3.3 | Probability Models

We assume that the models for the CMC scores follow a binomial distribution. These include

1. $Y^s \mid \theta_M \sim \text{Bin}(t_{nc}, \theta_M)$: Model for CMC score resulting from comparing the crime scene cartridge cases, given that it is a match.
2. $Y^s \mid \theta_N \sim \text{Bin}(t_{nc}, \theta_N)$: Model for CMC score resulting from comparing the crime scene cartridge caes, given that it is a non-match.
3. $Y_i^M \mid \theta_M \overset{\text{ind}}{\sim} \text{Bin}(t_{nc}, \theta_M),\ \ i = 1, \dots, t_{nm}$: Model for CMC scores which are matches in the database.
4. $Y_i^N \mid \theta_N \overset{\text{ind}}{\sim} \text{Bin}(t_{nc}, \theta_N),\ \ i = 1, \dots, t_{nn}$: Model for CMC scores which are non-matches in the database.

$Bin(t_{nc}, \theta_M)$ is a binomial distribution with t_{nc} cells and probability of match θ_M. $Bin(t_{nc}, \theta_N)$ is also a binomial distribution with t_{nc} cells and probability of non-match θ_N. θ_M and θ_N characterize the success probability of obtaining a match and non-match, respectively, when a cell on the reference cartridge scan is compared to that of the target cartridge scan. The t_{nc} cells result from partitioning of the breech-face impression scan of the cartridge case from the first crime scene into a grid as in Figure 2 , each cell compared to a region on the breech-face scan of the cartridge cases from the second crime scene. We used 64 cells from partitioning cartridge case scan into 8×8 grid. t_{nm} is the total number of CMC scores that are matches in the database, while t_{nn} is the total number of CMC scores that are non-matches in the database.

3.4 | Two Stage Approach

The Two Stage Approach, as illustrated by Parker (1966), utilizes a comparison metric (Δ), which inputs two items and provides a value indicating the strength of association. If Δ is large when items are similar, it is called a *similarity metric*. Otherwise, if Δ is large when items are different, it is called a *dissimilarity metric*. As the name implies, the approach consists of two parts: a comparison stage and a significance stage.

The common source comparison stage involves testing whether the two items come from the same source. Suppose Δ is a similarity metric , τ is an appropriate threshold, and $CS-a$ and $CS-b$ are the items;

1. If $\Delta(CS - a, CS - b) \leq \tau$ then exclude and done.
2. If $\Delta(CS - a, CS - b) > \tau$ then the items are associated and move to stage 2.

When we associate the items, then there are two possibilities, which include true association (failed exclusion was correct) and coincidental association (failed exclusion was incorrect). Coincidental probability is used to assess stage 2. Small coincidental probability means true association and vice versa. Parker(1966) defines the coincidental probability as

$$CP = P(\text{ fail to exclude } \mid H_d)$$

3.4.1 | Comparison Stage

This stage compares the first crime scene cartridge case and the cartridge cases from the second crime scene using an appropriate comparison metric and threshold. We used the CMC algorithm and score as the comparison metric and output of the comparison metric, respectively. Song (2013) and Song (2018) suggested six as the threshold of the CMC scores. This threshold discretizes the space of CMC scores into values corresponding to match and non-match. When the CMC score from comparing the crime scene cartridge cases is less than or equal to 6, it is declared a non-match. Otherwise, it is stated as a match. Specifically, if

1. $\Delta \leq 6$, then we conclude that the two cartridge cases being compared do not share a common source.
2. $\Delta > 6$, then the two cartridge cases are associated and move to Stage 2.

3.4.2 | Significance Stage

When the CMC score is greater than 6, then there is evidential support that the two cartridge cases share a common source. From that, the number of non-matches among the forty cartridge case comparisons in the database is used to calculate the coincidental probability. The coincidental probability is the probability that the set of Non-matches in the database would be found to be a match according to the threshold. The coincidental probability can be defined as

$$CP = \frac{1}{t_{nn}} \sum_{i=1}^{t_{nn}} I[y_i^N > 6]$$

Small coincidental probability represents strong evidence that the cartridge case from the first crime scene and the cartridge cases from the second crime scene share a common source. Otherwise, it represents weak evidence.

3.5 | Score-based Likelihood Ratios

SLR is used to quantify the similarity between two items. It is based on the output of the comparison metric. The output of the comparison metric was modeled under two different propositions. The CMC algorithm is the comparison metric, and the output of the comparison metric is the CMC scores. This means, $\Delta(C - a, C - b) = y$ and for the example using the crime scene data then $\Delta(\text{CS-1}, \text{CS-2}i) = y^s$. We modeled the CMC scores under matches and non-matches in the database. To compute the SLR, we plugged the CMC score between the crime scene cartridge cases in the matches and the non-matches likelihood ratio. This allowed us to determine the strength of evidence that the crime scene cartridge cases are from the same source (firearm). The SLR for the common source scenario described, denoted by δ, is computed as

$$\delta(y^s) = \frac{\binom{t_{nc}}{y^s}\theta_M^{y^s}(1-\theta_M)^{t_{nc}-y^s}}{\binom{t_{nc}}{y^s}\theta_N^{y^s}(1-\theta_N)^{t_{nc}-y^s}},$$

3.5.1 | Estimation of Parameters

Maximum likelihood estimation was used to estimate the common source SLR parameters.

For matches (numerator density), the estimation data is the CMC scores of all pairwise comparisons of cartridge cases from the same gun in the database. We have t_{nm} CMC scores as the estimation data.

For non-matches (denominator density), the estimation data is the CMC scores of all pairwise comparisons of cartridge cases from different guns in the database. Here, we have t_{nn} CMC scores as the estimation data. The estimates of the parameters are

$$\hat{\theta}_M = \frac{\sum_{i=1}^{t_{nm}} y_i^M}{t_{nm} t_{nc}}$$
$$\hat{\theta}_N = \frac{\sum_{i=1}^{t_{nn}} y_i^N}{t_{nn} t_{nc}}$$

3.5.2 | Score-based Likelihood Ratio Scale

The Likelihood Ratio Scale, suggested by ENFSI (2015), provides information about what each SLR value indicates (see Table 3). The verbal scale ranges from "do not support" to "extremely strong support". When the denominator density is greater than the numerator density, $1/(\delta(y^s))$ can be used.

TABLE 3 Interpretations of values of the Score-based Likelihood Ratio. Each entry corresponds to a particular strength of evidence. The values also give an indication as to the extent of support for the defense and prosecution propositions.

Values of likelihood ratio	**Verbal scale**
1	The findings do not support one proposition over the other.
2-10	Weak support
10-100	Moderate support
100-1000	Moderately strong support
1000-10,000	Strong support
10,000-1,000,000	Very strong
1,000,000 and above	Extremely strong support

When $\delta(y^s)$ is greater than one, it indicates that the CMC score between the crime scene cartridge cases is more similar to the CMC scores for matches in the database, than for non-matches. Values of $\delta(y^s)$ greater than one also indicate more support for the prosecution's proposition than for the defense's. Also, when $\delta(y^s)$ is less than one, we have more support for the defense's proposition than for the prosecution's.

3.6 | Bayesian Approach

Here, we assessed the strength of evidence using posterior model probabilities and Bayes factor. The posterior model probabilities involve posterior predictive distributions and the jury's prior probability that the crime scene cartridge cases share a common source (match).

3.6.1 | Estimation of Parameters

Parameter estimation for θ_M and θ_N involves assigning their prior distributions and constructing likelihoods for the CMC scores in the database. We utilized the prior and likelihoods to get posterior distributions for the parameters. Also, the prior distributions represent our belief about these parameters.

3.6.1.1 | Priors

Using Conjugate priors for θ_M and θ_N, we have ;

$$\theta_M \sim \text{ Be } (a = 1/2, b = 1/2)$$
$$\theta_N \sim \text{ Be } (c = 1/2, d = 1/2)$$

where Be $(1/2, 1/2)$ is a beta distribution with parameters $1/2$ and $1/2$. The parameters of the prior distributions are defined as: a is the prior successes of match, b is the prior failures of match, c is the prior successes of non-match, and d is the prior failures of non-match.

3.6.1.2 | Posterior Distributions

Posterior distributions for θ_M and θ_N provides information about their plausible values. Each of the posterior distribution follows a beta distribution (see Appendix 1 for the derivation). The two posterior distributions are

$$\theta_M \mid Y^M \sim \text{Be}\left(a + \sum_{i=1}^{t_{nm}} y_i^M, b + \sum_{i=1}^{t_{nm}} (t_{nc} - y_i^M)\right)$$

$$\theta_N \mid Y^N \sim \text{Be}\left(c + \sum_{i=1}^{t_{nn}} y_i^N, d + \sum_{i=1}^{t_{nn}} (t_{nc} - y_i^N)\right)$$

We also assume that θ_M and θ_N are independent. These posterior distributions for the parameters also describe their underlying uncertainty and distributions updated upon meeting the data. That is because the uncertainty when comparing regions of the first crime scene cartridge case to that of the second crime scene is vital in determining whether it is a match or non-match. Information about the uncertainty is depicted through their density plots.

3.6.2 | Model Comparison

The model comparison involves assessing the strength of evidence that the crime scene cartridge cases share a common source. The two models under consideration include match and non-match. Given the observed CMC scores from comparing the cartridge case from the first crime scene to the cartridge cases from the second crime scene and known information from the database, the jury can utilize posterior model probabilities to assess the models' strength of evidence.

3.6.2.1 | Posterior Model Probabilities

The jury's verdict, upon assessment of the strength of evidence that the crime scene cartridge cases were fired from the same gun makes use of the posterior model probabilities. This assessment is based on two propositions: the crime scene cartridge cases match, and the crime scene cartridge cases do not match. The posterior model probabilities for match and non-match include $P(M \mid Y^s = y^s, Y^D)$ and $P(N \mid Y^s = y^s, Y^D)$ which are used to assess the strength of evidence that the crime scene cartridge cases were fired from the same gun. We have already defined that t_{nc} is the total number of cells from partitioning a breech-face impression of a cartridge scan, Y^M is the set of matches in the database, Y^N is the set of non-matches in the database, and y^s is the CMC score when comparing the crime scene cartridge cases. Now, we incorporate them in the posterior model probabilities calculation. The posterior model probabilities can be calculated as

$$P(M \mid Y^s = y^s, Y^D) = \frac{P(Y^s = y^s \mid Y^M)P(M)}{P(Y^s = y^s \mid Y^M)P(M) + P(Y^s = y^s \mid Y^N)P(N)}$$

$$P(N \mid Y^s = y^s, Y^D) = \frac{P(Y^s = y^s \mid Y^N)P(N)}{P(Y^s = y^s \mid Y^M)P(M) + P(Y^s = y^s \mid Y^N)P(N)}$$

$$P(Y^s = y^s \mid Y^M) = \text{Be-Bin}\left(y^s; t_{nc}, a + \sum_{i=1}^{t_{nm}} y_i^M, b + \sum_{i=1}^{t_{nm}} (t_{nc} - y_i^M)\right)$$

$$P(Y^s = y^s \mid Y^N) = \text{Be-Bin}\left(y^s; t_{nc}, c + \sum_{i=1}^{t_{nn}} y_i^N, d + \sum_{i=1}^{t_{nn}} (t_{nc} - y_i^N)\right)$$

$\text{Be-Bin}\left(y^s; t_{nc}, a + \sum_{i=1}^{t_{nm}} y_i^M, b + \sum_{i=1}^{t_{nm}} (t_{nc} - y_i^M)\right)$ is a beta-binomial distribution evaluated at y^s with parameters $t_{nc}, a + \sum_{i=1}^{t_{nm}} y_i^M$, and $b + \sum_{i=1}^{t_{nm}} (t_{nc} - y_i^M)$. $P(Y^s = y^s \mid Y^M)$ and $P(Y^s = y^s \mid Y^N)$ represent posterior predictive distributions (see Appendix 2 for the derivation). They provide information about possible outcomes of CMC scores given the observed data. Distributions of future observations can inform the decisions of the jury. $P(M)$ and $P(N)$ are probabilities representing a jury's belief of obtaining a match or non-match given any two cartridge cases. Since the jury's belief cannot be guessed, values ranging from 0 to 1 are used. $P(M)$ and $P(N)$ sum to 1, that is $P(M) = 1 - P(N)$. Additionally, we put the jury into three categories: unbiased, optimistic, and pessimistic. An unbiased jury has equal prior probability for both propositions (P(M)=0.5 and P(N)=0.5). An optimistic jury has a prior probability that the cartridge cases do not match to be greater than the prior probability that the cartridge cases match (P(M) ∈ [0,0.5), P(N) ∈ (0.5,1]). Lastly, a pessimistic jury, on the other hand, has a prior probability that the cartridge cases match to be greater than the prior probability that they do not match (P(M) ∈ (0.5,1], P(N) ∈ [0,0.5)).

3.6.2.2 | Bayes factor

Bayes factor compares the integrated likelihood function (since the parameters θ_N and θ_N have been integrated out, see Appendix 2) of a CMC score under the two propositions (defense and prosecution). The integrated likelihood functions are posterior predictives (Ommen and Saunders, 2021, Equation (23)). It provides a metric for measuring the strength of evidence, whether the two cartridge cases match or do not match. The Bayes factor is computed as

$$\begin{aligned} BF(M : N) &= \frac{P(y^s|M)}{P(y^s|N)} \\ &= \frac{\text{Be-Bin}\left(y^s; t_{nc}, a + \sum_{i=1}^{t_{nm}} y_i^M, b + \sum_{i=1}^{t_{nm}} (t_{nc} - y_i^M)\right)}{\text{Be-Bin}\left(y^s; t_{nc}, c + \sum_{i=1}^{t_{nn}} y_i^N, d + \sum_{i=1}^{t_{nn}} (t_{nc} - y_i^N)\right)} \end{aligned}$$

A Bayes factor greater than one implies the CMC score is more likely under the matches model, while the CMC score is more likely under the non-matches model when the Bayes factor is less than one. Table 3 also provides a verbal scale for interpreting the Bayes factor. $\frac{1}{BF(M:N)}$ can also be used when $P(y^s|N)$ is greater than $P(y^s|M)$ because of the range of values in Table 3 .

3.7 | Leave-one-barrel-out Cross-validation

Comparing the performance of the methods deals with evaluating and assessing the effectiveness of using the three methods (two-stage approach, SLR, Bayesian approach) to measure the strength of evidence that a cartridge case from a crime scene matches a cartridge case from a different crime scene. Evaluation metrics, including sensitivity, specificity, and accuracy, compare performance. Since each method for assessing the strength of evidence has a defined path, the need to compare their performance aid informed decisions, error evaluation, resource utilization, and benchmarks.

We utilized the database for performance comparison. That is because the database contains the ground truth of each cartridge case pair comparison. Leave-one-barrel-out cross-validation technique is employed on the database. Removing a barrel entails removing that barrel from the database, and the CMC scores are used as the test set. The remaining CMC scores are used as the training set. Specifically, we utilize 10-fold cross-validation since the cartridge cases in the database are from ten barrels. Each fold in an iteration contains CMC scores that are matches or non-matches. Some of the CMC scores in the test set of an iteration would also be among the test set of another iteration due to the pairwise comparison of the cartridge cases. This results in a certain level of dependency among the test sets. Since there are ten barrels, the cross-validation iteration is done ten times. So final values of the evaluation metrics result from averaging values from the ten iterations. The cross-validation is also illustrated systematically as;

1. Let $b = \{b_1, b_2, \dots, b_{10}\}$ represent a dataset where each b_i $(i = 1, 2, \dots, 10)$ contains CMC scores resulting from comparing cartridges from b_i to any other cartridge (including comparing cartridges from b_i).
2. $\{b_1, b_2, \dots, b_{10}\}$ make up 10 folds. Each fold is used as a test set, while the remaining folds are used as the train set.
3. The method is trained on each $b - b_i$ and evaluated on b_i.
4. Performance metrics (sensitivity, specificity, accuracy) are collected for each fold. The average performance across all ten folds is computed to estimate the methods' performance.

3.7.1 | Making Decision

Since the CMC scores in the database are already a set of matches and non-matches, it is necessary to determine the decisions of each of the three methods on the test set. Whether match or non-match, the decisions allow the computation of the performance metrics. So, we compare the decisions of the methods on the test to their ground truth. The decision is depicted as

$$\text{d} = \begin{cases} \text{Match }, & \text{if } dc > t \\ \text{Non-match }, & \text{if } dc \leq t \end{cases}$$

where dc is the decision criteria and t is the threshold.

3.7.1.1 | Two-stage Approach

We used stage 1 of the two-stage approach to decide whether a CMC score between a pair of cartridges in a test set is a match or non-match. As stated in Section 3.4.1, Song(2013) suggests six as the CMC threshold. Thus, a CMC score greater than 6 is a match. Otherwise, it is a non-match. We have that

$$\text{d} = \begin{cases} \text{Match}, & \text{if } y > 6 \\ \text{Non-match}, & \text{if } y \leq 6 \end{cases}$$

Here, y is a CMC score between a pair of cartridge case in a test set and 6 is the decision threshold.

3.7.1.2 | SLR

For each iteration of the cross-validation, the train set is used to estimate the parameters of the SLR. The set of matches in a train set is used to estimate the parameters of the numerator density. In contrast, the parameters of the denominator density are estimated using the set of non-matches. We plug CMC scores from test sets into the SLRs to determine whether it is a match or a non-match. When $SLR > 1$, the CMC score is classified as a match. Otherwise, it is a non-match, as shown below

$$\text{d} = \begin{cases} \text{Match}, & \text{if } \delta(b_i) > 1 \\ \text{Non-match}, & \text{if } \delta(b_i) \leq 1 \end{cases}$$

where $\delta(b_i)$ is the SLR value for comparing cartridge cases involving barrel b_i, and 1 is the threshold. When $log(\delta(b_i))$ is used, positive values are classified as a match, and negative values are classified as a non-match. The cross-validation SLR is computed as

$$\delta(b_i) = \frac{\binom{t_{nc}}{y}\theta_M^y(1-\theta_M)^{t_{nc}-y}}{\binom{t_{nc}}{y}\theta_N^y(1-\theta_N)^{t_{nc}-y}}, \; i = 1, 2, \ldots, 10$$

where

$$\hat{\theta}_M = \frac{\sum_{i=1}^{t_{nm}-b_i} y_i^M}{(t_{nm} - b_i)t_{nc}}$$
$$\hat{\theta}_N = \frac{\sum_{i=1}^{t_{nn}-b_i} y_i^N}{(t_{nn} - b_i)t_{nc}}$$

3.7.1.3 | Bayesian Approach

We utilized Bayes factors to make decisions in the Bayesian approach. Bayes factor compares the posterior predictive distribution of the CMC scores under matches and non-matches. Suppose $t_{nm} - b_i$ and $t_{nn} - b_i$ are the total numbers of matches and non-matches, respectively, in the train set of an iteration in the cross-validation; we then define the Bayes factor as

$$BF(M:N) = \frac{\text{Be-Bin}\left(y; t_{nc}, a + \sum_i^{t_{nm}-b_i} y_i^M, b + \sum_i^{t_{nm}-b_i}(t_{nc} - y_i^M)\right)}{\text{Be-Bin}\left(y; t_{nc}, c + \sum_i^{t_{nn}-b_i} y_i^N, d + \sum_i^{t_{nn}-b_i}(t_{nc} - y_i^N)\right)}$$

where Be-Bin $\left(y; t_{nc}, a + \sum_i^{t_{nm}-b_i} y_i^M, b + \sum_i^{t_{nm}-b_i}(t_{nc} - y_i^M)\right)$ is a beta-binomial distribution evaluated at y, and the rest are its parameters. In an iteration, parameters of the posterior predictive distribution for matches are estimated using the set of matches in the train set, while those of the non-matches are estimated using the set of non-matches. y represents each of the CMC scores in the test of an iteration. When $BF(M:N) > 1$, the particular test set CMC score is classified as a match. Otherwise, it is classified as a non-match. We have that

$$\text{d} = \begin{cases} \text{Match}, & \text{if } BF(M:N) > 1 \\ \text{Non-match}, & \text{if } BF(M:N) \leq 1 \end{cases}$$

When $log(BF(M:N))$ is used, positive values are classified as a match, and negative values are classified as a non-match

3.7.2 | Performance Metrics

The following definitions were used to compute the performance metrics and are based on a test set:

1. P (TP+FN) : The total number of matches = $\sum I(gt = \text{match})$, where gt is the ground truth.
2. N (TN+FP) : The total number of non-matches = $\sum I(gt = \text{non-match})$
3. T (P+N=TP+TN+FP+FN) : The total number of CMC scores or observations in a test set = $\sum I(gt = \text{match}) + \sum I(gt = \text{non-match})$
4. True Positive (TP) : The number of correctly identified matches = $\frac{\sum I(d=\text{match})}{\sum I(gt=\text{match})}$
5. True Negative (TN) : The number of correctly identified non-matches = $\frac{\sum I(d=\text{non-match})}{\sum I(gt=\text{non-match})}$
6. False Positive (FP) : The number of non-matches incorrectly identified as matches = $1 - \frac{\sum I(d=\text{non-match})}{\sum I(gt=\text{non-match})}$
7. False Negative (FN) : The number of matches incorrectly identified as non-matches = $1 - \frac{\sum I(d=\text{match})}{\sum I(gt=\text{match})}$

3.7.2.1 | Sensitivity and Specificity

Sensitivity is the proportion of matches with supporting results.

$$Sensitivity = \frac{TP}{TP + FN}$$

Specificity is the proportion of non-matches with supporting results.

$$Specificity = \frac{TN}{TN + FP}$$

3.7.2.2 | Accuracy

Accuracy is the closeness of the result to the correct result. It measures the total number of correct decisions among the total number of comparisons.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

We express accuracy also as a weighted average of sensitivity and specificity. We have that;

$$Accuracy = \frac{P}{T} \cdot \frac{TP}{P} + \frac{N}{T} \cdot \frac{TN}{N} = \frac{P}{T} \cdot \text{ sensitivity} + \frac{N}{T} \cdot \text{ specificity}$$

4 | RESULTS

This section presents the decisions about the strength of evidence based on the two-stage approach, score-based likelihood ratio, and Bayesian approach. We also show the performance of each of the methods.

4.1 | Two Stage Approach

Table 4 shows the decision of the comparison stage (first stage) of the two-stage approach. We compared CS-1 to CS-2a and CS-2b and their decisions provided. When comparing the cartridge case from the first crime scene (CS-1) to the first cartridge case from the second crime scene (CS-2a), the CMC obtained (which is 3) is less than the threshold (6), while when compared to the second cartridge case from the second crime scene (CS-2b), the CMC obtained (which is 7) is above the threshold. This implies that we do not have an association at the comparison stage or that there is support for the defense's proposition when CS-1 and CS-2a are compared. In contrast, we have an association or support for the prosecution's proposition when CS-1 and CS-2b are compared. So, for the first case (CS-1 and CS-2a), we exclude and do not move on to stage 2 (significance stage) of the two-stage approach. However, for the second case (CS-1 and CS-2b), we move on to stage 2 since the CMC score implies support for the prosecution's proposition. Now, the coincidental probability equals 0 since no non-match exists in the cartridge case scans database with a CMC score greater than 6. The value for the coincidental probability indicates strong support that CS-1 and CS-2b were fired from the same firearm.

TABLE 4 Comparison stage of the two-stage approach showing whether the CMC scores between the compared cartridge cases from the two different crime scenes are greater than the threshold. That represents stage 1 of the two-stage approach. The threshold partitions the scores into support for the prosecution's proposition and vice versa.

Cartridge Cases	CMC > 6
CS-1 and CS-2a	False
CS-1 and CS-2b	True

4.2 | SLR

We computed SLR values for the two cases. Table 5 depicts the SLR and log(SLR) values. Either log(SLR) or SLR can be used to measure the strength of evidence that the crime scene cartridge cases were fired from the same firearm in the SLR method. Table 3 shows that the SLR value when CS-1 and CS-2a are compared indicates moderate support for the defense's proposition. This stems from using $\frac{1}{SLR}\left(\frac{1}{0.019865}=50.3397\right)$ as the SLR value instead of 0.019865, since 0.019865 is less than one and cannot be referred from Table 3 . However, when CS-1 and CS-2b are compared, the SLR value implies very strong support for the prosecution's proposition. The implication of the log(SLR) values is equivalent to their SLR interpretations.

TABLE 5 SLR and log(SLR) values measuring the strength of evidence that the cartridge cases from the two different crime scenes were fired from the same firearm. The values gives an indication of the extent of support for the defense and prosecution propositions.

Cartridge Cases	SLR	log(SLR)
CS-1 and CS-2a	0.019865	-3.918756
CS-1 and CS-2b	98284.54	11.495622

4.3 | Bayesian Approach

Figure 4 shows the posterior model probabilities for the two cases. The x-axis shows the jury's prior probability that the cartridge cases match, and the y-axis also shows the posterior model probabilities that they match. In Figure 4 , the posterior model probability is an increasing or decreasing function of the jury's prior probability. The value of the posterior model probability when comparing CS-1 and CS-2b is equal to 1, except when the jury's prior probability that they match is zero (P(M)=0). This implies that we have strong evidence that CS-1 and CS-2b were fired from the same firearm if and only if the jury's prior probability that they match is not zero. When comparing CS-1 and CS-2a, the posterior model probability that they match is less than 0.50 when the jury's prior probability that they match is less than 0.96. Thus, we have strong evidence that CS-1 and CS-2a match if the jury's prior belief that they match is greater than or equal to 0.96, while we have strong evidence that CS-1 and CS-2a do not match if the jury's prior probability that the cartridge cases match is less than 0.96. We then define the posterior model probabilities for the three jury categories mentioned in Section 3.6.2.1 as follows:

1. Unbiased: $P(M|3, Y^D) = 0.02$, $P(M|7, Y^D) = 1.00$
2. Optimistic: $P(M|3, Y^D) \in [0, 0.02)$, $P(M|7, Y^D) = 1.00$ when $P(M) \neq 0$
3. Pessimistic: $P(M|3, Y^D) \in (0.02, 1]$, $P(M|7, Y^D) = 1.00$

In the Bayesian approach, we also used the Bayes factor to measure the strength of evidence. Table 6 depicts the two cases' Bayes factor and log(Bayes factor) values. Table 3 helps us to interpret the Bayes factor values. It can be seen from Table 6 that the Bayes factor when comparing CS-1 and CS-2a is less than 1, so it is appropriate to use $\frac{1}{\text{Bayes factor}}\left(\frac{1}{0.0212509}=47.0568\right)$ as the Bayes factor value so that we can refer from Table 3 . Now, with 47.0568, we have moderate evidence that CS-1 and CS-2a were not fired from the same firearm (moderate support for the defense's proposition). Regarding the Bayes factor when comparing CS-1 and CS-2b, it indicates very strong evidence that they were fired from the same firearm. Here, interpretations of the log(Bayes factor) values are also equivalent to their Bayes factor values.

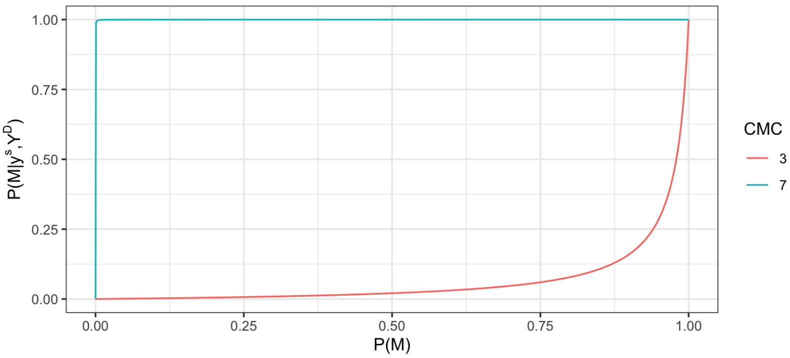

FIGURE 4 Posterior model probabilities measuring the strength of evidence that the cartridge cases from two different crime scenes were fired from the same firearm. The plot shows the posterior probability of a match and the jury's prior probability of a match. Large values of prior probability of a match correspond to large values of posterior model probabilities of a match.

TABLE 6 Bayes factor and log(Bayes factor) values measuring the strength of evidence that the cartridge cases from the two different crime scenes were fired from the same firearm. The values give an indication of the extent of support for the defense and prosecution propositions.

Cartridge Cases	Bayes Factor	log(Bayes Factor)
CS-1 and CS-2a	0.0212509	-3.851352
CS-1 and CS-2b	93302.29	11.4436

4.4 | Leave-one-barrel-out Cross-validation

Table 7 shows a summary of the number of observations (CMC scores) in the leave-one-barrel-out cross-validation. The number of cartridge cases produced by the barrels ranges from 3 to 5. Barrels labelled as 1, 4, and 9 have the same number of cartridge cases. Barrels labelled as 2, 5, 6, and 10 also have the same number of cartridge cases. Lastly, barrels labelled as 3, 7, and 8 also have the same number of cartridge cases. From Table 7 , barrels with the same number of cartridge cases have the same number of matches and non-matches in their test sets and train sets. The number of CMC scores in the test sets and train sets range from 114 to 185 and 1312 to 1383, respectively. Moreover, the total number of cartridge cases (which is 40) and the total number of matches in Table 7 (which is 63) add up to that of the processed cartridge cases scans database described in Section 2.2. However, the total number of non-matches in Table7 (which is 1434) does not add up to the actual number of non-matches in the processed cartridge cases scans database (which is 717) (see Section 2.2). That is due to multiple counts of scores from pairwise comparisons of cartridge cases from different barrels.

In determining the performance of the approaches for assessing the strength of evidence, we employed leave-one-barrel-out cross-validation on the database. Figure 5 depicts the boxplots for all matches and non-matches from the cross-validation based on the two-stage approach, SLR, and Bayes factor. The x-axis of each subfigure is the ground truth, and it shows the actual class of the CMC scores from a pairwise comparison of cartridge cases in the database. The points' locations are the predicted class (match or non-match) in the various test sets of the cross-validation. The y-axis of the two-stage approach is the CMC scores, which is the output of the similarity metric in stage 1 of the two-stage approach. The y-axis of the boxplot for the score-based likelihood ratio and Bayes factor are the SLR and Bayes factor values, respectively. The horizontal line in each subfigure is the decision threshold for prediction. We used 6, 1, and 1 as decision thresholds for the two-stage approach, SLR and Bayes factor, respectively. The CMC scores that are non-matches are all predicted correctly (below the thresholds). However, some scores are matches based on their ground truth, but they are predicted to be non-matches (below the thresholds). The points are differentiated by ten barrels, making up ten test sets in the cross-validation.

Figure 6 and Table 8 show the performance metrics' average proportions from the cross-validation. The SLR and Bayes factor have equal values. That implies that they have the same performance for assessing the strength of evidence that the crime

TABLE 7 Number of cartridge cases, matches, and non-matches pertaining to each barrel in the leave-one-out cross-validation. The ground truths of the CMC scores in each test set are used to evaluate the performance of the two-stage approach, Score-based Likelihood ratio, and Bayes factor.

Barrel Label (fold)	Number of cartridge cases	Number of matches in train set	Number of non-matches in train set	Number of matches in test set	Number of non-matches in test set
1	3	60	1323	3	111
2	4	57	1290	6	144
3	5	53	1259	10	175
4	3	60	1323	3	111
5	4	57	1290	6	144
6	4	57	1290	6	144
7	5	53	1259	10	175
8	5	53	1259	10	175
9	3	60	1323	3	111
10	4	57	1290	6	144

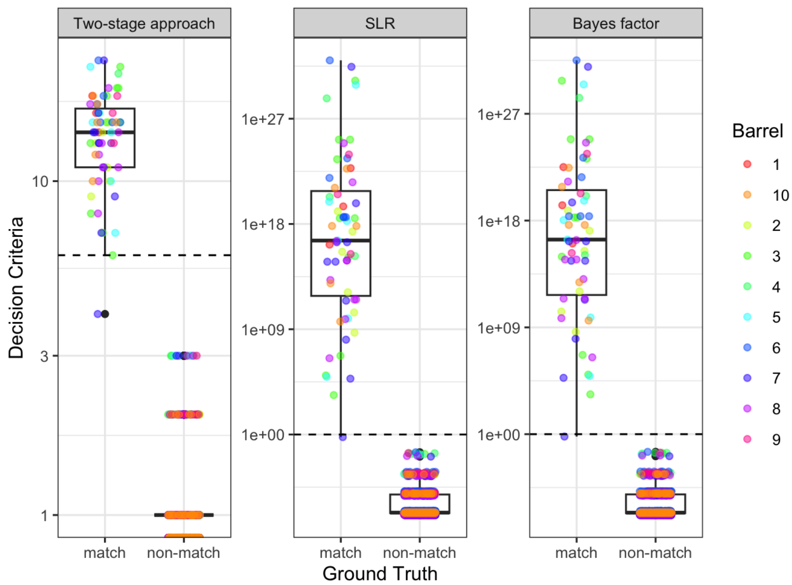

FIGURE 5 Boxplots for all matches and non-matches from the leave-one-barrel-out cross-validation generated by the two-stage approach, SLR, and Bayes factor. The points are differentiated by the ten barrels. The ten barrels represent ten folds in the cross-validation. So, each point represents the result of an observation in a particular barrel. The ground truth (whether match or non-match) is compared to the methods' (two-stage approach, SLR, Bayes factor) predicted decision. The decision thresholds are indicated by the horizontal lines.

scene cartridge cases were fired from the same firearm. The two-stage approach has slightly distinct values for sensitivity and accuracy.

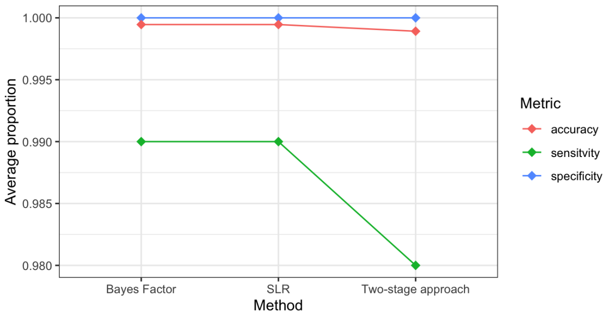

FIGURE 6 Parallel coordinate plot of average proportions of the performance metrics including sensitivity, specificity, and accuracy from the leave-one-barrel-out cross-validation. Each line represents the average proportions of the metric for the three methods. The values of the metrics range from 0.98 to 1.00.

TABLE 8 Average proportions of the performance metrics including sensitivity, specificity, and accuracy from the leave-one-barrel-out cross-validation shown as a contingency table. Each row is the average proportions of the metrics for the particular method.

	Metric		
Method	**Sensitivity**	**Specificity**	**Accuracy**
Two stage approach	0.980	1.00	0.998
SLR	0.990	1.00	0.999
Bayes factor	0.990	1.00	0.999

5 | DISCUSSION

This study assessed the strength of evidence that cartridge cases from two different crime scenes were fired from the same firearm. We also compared the performance of using the two-stage approach, SLR and Bayes factor, to assess the strength of evidence. For the two cases (CS-1 and CS-2a, CS-1 and CS-2b) under consideration, the three methods provided almost similar conclusions when assessing the strength of evidence. Leave-one-barrel-out (10-fold) cross-validation was employed on the database to evaluate the performance of the methods, since the database contains the ground truth of each pairwise comparison. It was revealed that the score-based likelihood ratio and Bayes factor have equal performance according to the metrics (sensitivity, specificity, accuracy). Additionally, it is also evident that all the methods performed well so far as we obtained values above 98% for the metrics. The threshold of the two-stage approach plays a crucial role in determining whether we can move on to stage 2. Varying the threshold can improve the performance metrics of the two stage approach compared to the SLR and Bayesian approach. A small coincidental probability for case 2 (CS-1 and CS-2b) provides proof and substantiates the support for the prosecution's proposition. Using the binomial distribution, the SLR values for case 1 and case 2 yielded moderate and very strong support for the defense's and prosecutor's propositions, respectively. That also substantiates the results from the two-stage approach. The SLR and Bayes factor are each based on ratios of the propositions using the CMC scores. For these two ratio-based approaches, the jury can utilize either the SLR or Bayes factor to inform their decision as to whether the crime scene cartridge cases were fired from the same firearm. The posterior model probabilities measure the strength of evidence considering the jury's initial belief. With the jury's prior probability of a match, a high prior probability of a match eventually attains a strong evidence of a match. Since the computation of the posterior model probabilities already includes the jury's prior belief, it gives the jury an easy metric to make a verdict. We can evaluate and assess the verdict by computing the posterior model probabilities for each corresponding value of the jury's prior belief.

Song(2018) used cumulative false positive error rate and cumulative false negative error rate to assess error rates of firearm identification (using the database). The cumulative false positive error rate was based on the sum of probabilities for CMC scores between 6 and 26, estimated by the distribution for non-matches. The cumulative false negative error rate was also based on the sum of the probabilities for CMC scores between 0 and 5, estimated by the distribution for matches. 6.1×10^{-10} and 2.1×10^{-3} were obtained as values for the cumulative false positive error rate and cumulative false negative error rate, respectively. Even so, it contrasts with this study in which we employed leave-one-barrel-out cross-validation with the two-stage approach, SLR and Bayes factor, to assess the error rates.

The common-source scenario was employed in this study. That was due to the size of the example data and database. The example data assumed one cartridge case from one crime scene and two cartridge cases from another, coupled with at most five cartridge cases from ten barrels in the database. The specific source scenario, as illustrated in Ommen and Saunders (2021), will be based on assessing whether a cartridge case from a crime scene was fired from a suspect's gun. Here, enough cartridge cases from the suspect's gun are necessary for the specific source analysis. The comparison stage of the two-stage approach is similar in both scenarios. The significance stage of the two-stage approach for the common-source scenario deals with all pairwise comparisons of cartridge cases from different barrels (source) in the database according to the similarity metric. On the other hand, the significance stage for the specific-source scenario has three approaches, which include trace-anchored (comparing the cartridge case from the crime scene to all cartridge cases in the database), source-anchored (comparing the cartridge case from the suspect's gun to all cartridge cases in the database), and general-match (comparing cartridge cases from different barrels in the database). Now, it seems one cartridge case from the suspect's gun is enough for the two-stage approach application. The SLR is based on the ratio of two densities (that of the defense and prosecution). The numerator density of the common source SLR is estimated using all CMC scores from comparing cartridge cases from the same barrel in the database. In contrast, the denominator density of the common-source SLR is estimated using all CMC scores from comparing cartridge cases from different barrels in the database. The specific source SLR includes fully trace-anchored (the numerator density is estimated using CMC scores from comparing the cartridge case from the crime scene to other cartridge cases from the suspect's barrel and the denominator density is estimated using CMC scores from comparing the cartridge case from the crime scene to the cartridge cases in the database), fully source-anchored (the numerator density is estimated using CMC scores from comparing the cartridge case from the suspect's barrel to other cartridge cases from the suspect's barrel and the denominator density is estimated using CMC scores from comparing the cartridge case from the suspect's barrel to all cartridge cases in the database), asymmetric-anchored (the numerator density is estimated using CMC scores from comparing the cartridge case from the suspect's barrel to other cartridge cases from the suspect's barrel and the denominator density is estimated using CMC scores from comparing the cartridge case from the crime scene to the cartridge cases in the database), and general match (the numerator density is estimated using CMC scores from comparing the cartridge case from the suspect's barrel to other cartridge cases from the suspect's barrel and the denominator density is estimated using CMC scores from comparing cartridge cases from different barrels in the database). Each has a different estimation procedure for the numerator and denominator density, and even difficult to apply without enough cartridge cases from the suspect's barrel. The posterior model probability and Bayes factor face similar difficulty.

The forty cartridge cases in the database were fired from ten Ruger 9 mm barrels. This adds to the high clarity and discriminatory power of the distribution of matches and non-matches, since few cartridge cases were obtained consecutively from each barrel. More cartridge cases from each barrel will lead to a complete characterization of the distribution of CMC scores from pairwise comparison of cartridge cases from the same barrel. The number of non-matches will tend to increase excessively as the number of barrels and firearm brands increase, Thus, a CMC score in the database has high probability of being classified as a non-match when there is an increase of the number barrels and firearm brands.

The two-stage approach suffers from varying thresholds for the CMC scores at the comparison stage. For instance, a threshold of seven leads to support for the defense proposition for the two cases and also changes the values of its performance metrics. Given different threshold values, there is a tendency of getting different conclusions at the comparison stage. We used the binomial distribution to model the CMC scores for the SLR. Nonetheless, it is possible to use different distributions, such as beta-binomial distribution. That might lead to similar or different conclusions for the SLR approach. Partition of the CMC scores in the database into ten folds using the ten barrels leads to CMC scores in multiple folds. That is due to a comparison of cartridge cases from different barrels. Specifically, CMC scores that are matches do not have duplicate observations in a different fold. On the other hand, those that are non-matches are duplicated in the folds (barrels) that the compared cartridge cases emanate. So, there are repeated CMC scores in the non-matches class.

This study underscores the efficacy of using statistical methods to assess the strength of evidence obtained from breech-face impressions on cartridge cases. That is substantiated by the values of the metrics. Thus, applying statistical methods in firearm legal proceedings adds to the resulting justice level and fairness.

6 | APPENDIX

1. **Posterior Distributions**
   $Y_i^M \mid \theta_M \overset{\text{ind}}{\sim} \text{Bin}(t_{nc}, \theta_M),\ \ i = 1, \ldots, t_{nm}$ is the model for CMC scores that are matches. $Y_i^N \mid \theta_N \overset{\text{ind}}{\sim} \text{Bin}(t_{nc}, \theta_N),\ \ i = 1, \ldots, t_{nn}$ is also the model for CMC scores that are non-matches. Bin is a binomial distribution.
   Prior distributions for θ_M and θ_N are respectively Be (a, b) and Be (c, d). Be is a beta distribution.

   (a) **Posterior distribution for θ_M**

$$
\begin{aligned}
P(\theta_M \mid Y^M) &\propto P(Y^M \mid \theta_M) P(\theta_M) \\
&= \left[\prod_{i=1}^{t_{nm}}\right] \binom{t_{nc}}{y_i^M} \theta_M^{\sum_{i=1}^{t_{nm}} y_i^M} (1-\theta_M)^{\sum_{i=1}^{t_{nm}} (t_{nc} - y_i^M)} \frac{1}{B(a,b)} \theta_M^{a-1} (1-\theta_M)^{b-1} \\
&\propto \theta_M^{a + \sum_{i=1}^{t_{nm}} y_i^M - 1} (1-\theta_M)^{b + \sum_{i=1}^{t_{nm}} (t_{nc} - y_i^M) - 1} \\
&\Rightarrow \theta_M \mid Y^M \sim Be\left(a + \sum_{i=1}^{t_{nm}} y_i^M, b + \sum_{i=1}^{t_{nm}} (t_{nc} - y_i^M) \right)
\end{aligned}
$$

   (b) **Posterior distribution for θ_N**

$$
\begin{aligned}
P(\theta_N \mid Y^N) &\propto P(Y^N \mid \theta_N) P(\theta_N) \\
&= \left[\prod_{i=1}^{t_{nn}}\right] \binom{t_{nc}}{y_i^N} \theta_N^{\sum_{i=1}^{t_{nn}} y_i^N} (1-\theta_N)^{\sum_{i=1}^{t_{nn}} (t_{nc} - y_i^N)} \frac{1}{B(c,d)} \theta_N^{c-1} (1-\theta_N)^{d-1} \\
&\propto \theta_N^{c + \sum_{i=1}^{t_{nn}} y_i^N - 1} (1-\theta_N)^{d + \sum_{i=1}^{t_{nn}} (t_{nc} - y_i^N) - 1} \\
&\Rightarrow \theta_N \mid Y^N \sim Be\left(c + \sum_{i=1}^{t_{nn}} y_i^N, d + \sum_{i=1}^{t_{nn}} (t_{nc} - y_i^N) \right)
\end{aligned}
$$

2. **Posterior Predictive Distributions**
   $Y^s \mid \theta_M \sim \text{Bin}(t_{nc}, \theta_M)$ is the model for new CMC score resulting from comparing the crime scene cartridge cases, given that it is a match. $Y^s \mid \theta_N \sim \text{Bin}(t_{nc}, \theta_N)$ is also the model for new CMC score resulting from comparing the crime scene cartridge cases, given that it is a non-match.

   (a) **Posterior predictive distribution conditional on matches**

$$
\begin{aligned}
P(Y^s = y^s \mid Y^M) &= \int_0^1 P(Y^s = y^s \mid \theta_M) P(\theta_M \mid Y^M) d\theta_M \\
&= \int_0^1 \binom{t_{nc}}{y^s} \theta_M^{y^s} (1-\theta_M)^{t_{nc} - y^s} \frac{\theta_M^{a + \sum_{i=1}^{t_{nm}} y_i^M - 1} (1-\theta_M)^{b + \sum_{i=1}^{t_{nm}} (t_{nc} - y_i^M) - 1}}{B(a + \sum_{i=1}^{t_{nm}} y_i^M, b + \sum_{i=1}^{t_{nm}} (t_{nc} - y_i^M))} d\theta_M \\
&= \frac{\binom{t_{nc}}{y^s} B(y^s + a + \sum_{i=1}^{t_{nm}} y_i^M, t_{nc} - y^s + b + \sum_{i=1}^{t_{nm}} (t_{nc} - y_i^M))}{B(a + \sum_{i=1}^{t_{nm}} y_i^M, b + \sum_{i=1}^{t_{nm}} (t_{nc} - y_i^M))} \int_0^1 \frac{\theta_M^{y^s + a + \sum_{i=1}^{t_{nm}} y_i^M - 1} (1-\theta_M)^{t_{nc} - y^s + b + \sum_{i=1}^{t_{nm}} (t_{nc} - y_i^M) - 1}}{B(y^s + a + \sum_{i=1}^{t_{nm}} y_i^M, t_{nc} - y^s + b + \sum_{i=1}^{t_{nm}} (t_{nc} - y_i^M))} \\
&\quad d\theta_M \\
&= \frac{\binom{t_{nc}}{y^s} B(y^s + a + \sum_{i=1}^{t_{nm}} y_i^M, t_{nc} - y^s + b + \sum_{i=1}^{t_{nm}} (t_{nc} - y_i^M))}{B(a + \sum_{i=1}^{t_{nm}} y_i^M, b + \sum_{i=1}^{t_{nm}} (t_{nc} - y_i^M))} \\
\therefore P(Y^s = y^s \mid Y^M) &= \text{Be-Bin}\left(y^s; t_{nc}, a + \sum_{i}^{t_{nm}} y_i^M, b + \sum_{i}^{t_{nm}} (t_{nc} - y_i^M) \right)
\end{aligned}
$$

(b) **Posterior predictive distribution conditional on non-matches**

$$
\begin{aligned}
P(Y^s = y^s \mid Y^N) &= \int_0^1 P(Y^s = y^s \mid \theta_N) P(\theta_N \mid Y^N) d\theta_N \\
&= \int_0^1 \binom{t_{nc}}{y^s} \theta_N^{y^s} (1-\theta_N)^{t_{nc}-y^s} \frac{\theta_N^{c+\sum_{i=1}^{t_{nn}} y_i^N - 1} (1-\theta_N)^{d+\sum_{i=1}^{t_{nn}}(t_{nc}-y_i^N)-1}}{B(c+\sum_{i=1}^{t_{nn}} y_i^N, d+\sum_{i=1}^{t_{nn}}(t_{nc}-y_i^N))} d\theta_N \\
&= \frac{\binom{t_{nc}}{y^s} B(y^s + c + \sum_{i=1}^{t_{nn}} y_i^N, t_{nc} - y^s + d + \sum_{i=1}^{t_{nn}}(t_{nc}-y_i^N))}{B(c+\sum_{i=1}^{t_{nn}} y_i^N, d+\sum_{i=1}^{t_{nn}}(t_{nc}-y_i^N))} \int_0^1 \frac{\theta_N^{y^s+c+\sum_{i=1}^{t_{nn}} y_i^N - 1} (1-\theta_N)^{t_{nc}-y^s+d+\sum_{i=1}^{t_{nn}}(t_{nc}-y_i^N)-1}}{B(y^s + c + \sum_{i=1}^{t_{nn}} y_i^N, t_{nc} - y^s + d + \sum_{i=1}^{t_{nn}}(t_{nc}-y_i^N))} \\
&\quad d\theta_N \\
&= \frac{\binom{t_{nc}}{y^s} B(y^s + c + \sum_{i=1}^{t_{nn}} y_i^N, t_{nc} - y^s + d + \sum_{i=1}^{t_{nn}}(t_{nc}-y_i^N))}{B(c+\sum_{i=1}^{t_{nn}} y_i^N, d+\sum_{i=1}^{t_{nn}}(t_{nc}-y_i^N))} \\
\therefore P(Y^s = y^s \mid Y^N) &= \text{Be-Bin}\left(y^s; t_{nc}, c + \sum_i^{t_{nn}} y_i^N, d + \sum_i^{t_{nn}} (t_{nc} - y_i^N) \right)
\end{aligned}
$$

Be-Bin is a beta-binomial distribution.

7 | ACKNOWLEDGEMENTS

8 | BIBLIOGRAPHY

References

1. Devine DJ. *Jury decision making: The state of the science*. NYU Press . 2012.

2. Song J. Proposed NIST ballistics identification system (NBIS) based on 3D topography measurements on correlation cells. *AFTE Journal* 2013; 45(2): 184–194.

3. Geradts ZJ, Bijhold J, Hermsen R, Murtagh F. Image matching algorithms for breech face marks and firing pins in a database of spent cartridge cases of firearms. *Forensic science international* 2001; 119(1): 97–106.

4. Zemmels J, VanderPlas S, Hofmann H. A Study in Reproducibility: The Congruent Matching Cells Algorithm and cmcR Package. 2022.

5. Tong M, Song J, Chu W, Thompson RM. Fired cartridge case identification using optical images and the congruent matching cells (CMC) method. *Journal of research of the National Institute of Standards and Technology* 2014; 119: 575.

6. Ommen DM, Saunders CP. A problem in forensic science highlighting the differences between the Bayes factor and likelihood ratio. *Statistical Science* 2021; 36(3): 344–359.

7. Parker J. A statistical treatment of identification problems. *Jornal of the Forensic Science Society* 1966; 6(1): pp. 33-39.

8. Song J, Vorburger TV, Chu W, et al. Estimating error rates for firearm evidence identifications in forensic science. *Forensic science international* 2018; 284: 15–32.

9. ENFSI Guideline for Evaluative Reporting in Forensic Science, European Network of Forensic Science Institutes, Dublin,. 2015: p. 17.