REPORT DOCUMENTATION PAGE	Form Approved OMB NO. 0704-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggesstions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA, 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any oenalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
24-10-2016	Final Report	1-Sep-2012 - 31-Jul-2015

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Final Report: Statistical Analysis of Firearms/Toolmarks Interpretation of cartridge case evidence using IBIS and Bayesian networks	W911NF-12-2-0056
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHORS	5d. PROJECT NUMBER
Keith B. Morris, Eric F. Law, Roger L. Jefferys, and Elizabeth C. Dearth	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES	8. PERFORMING ORGANIZATION REPORT NUMBER
West Virginia University Research Corporat 886 Chestnut Ridge Road P.O. Box 6845 Morgantown, WV 26506 -6845	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)	10. SPONSOR/MONITOR'S ACRONYM(S) ARO
U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211	11. SPONSOR/MONITOR'S REPORT NUMBER(S) 61989-LS-FRD.5

12. DISTRIBUTION AVAILIBILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT

The IBIS system provides a means of correlating the images of two breech face or firing pin impressions. Cartridges fired by the same gun result in similar images and thus higher scores. The generated scores, together with related firearm and ammunition information were transformed into a Bayesian network. Bayesian networks allow for the assessment of evidence based upon two propositions (same gun or different gun). This allows a forensic scientist to provide insight to courts and investigators as to the value of the evidence.

15. SUBJECT TERMS

Firearms statistical analysis Bayesian networks IBIS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Keith Morris
UU	UU	UU	UU		19b. TELEPHONE NUMBER 304-293-3169

Standard Form 298 (Rev 8/98)
Prescribed by ANSI Std. Z39.18

Report Title

Final Report: Statistical Analysis of Firearms/Toolmarks
Interpretation of cartridge case evidence using IBIS and Bayesian networks

ABSTRACT

The IBIS system provides a means of correlating the images of two breech face or firing pin impressions. Cartridges fired by the same gun result in similar images and thus higher scores. The generated scores, together with related firearm and ammunition information were transformed into a Bayesian network. Bayesian networks allow for the assessment of evidence based upon two propositions (same gun or different gun). This allows a forensic scientist to provide insight to courts and investigators as to the value of the evidence.
The breech face (BF) and firing pin (FP) scores, and their product, were used to assess the ability of the system to classify an "unknown" cartridge case into a same-gun or different-gun category. The IBIS system does not provide for an easy means to use the combination of the BF and FP scores. Twenty sets of known and questioned cartridge cases, from a large collection which had been analyzed by operational firearms examiners, were examined and tested using the Bayesian networks. Out of the 20 comparisons, there were eight true positives, seven true negatives, five false negatives, and zero false positives. In all instances of eliminations, the support for the different-gun hypothesis was, at minimum, strong.

Enter List of papers submitted or published that acknowledge ARO support from the start of the project to the date of this printing. List the papers, including journal references, in the following categories:

(a) Papers published in peer-reviewed journals (N/A for none)

Received Paper

TOTAL:

Number of Papers published in peer-reviewed journals:

(b) Papers published in non-peer-reviewed journals (N/A for none)

Received Paper

TOTAL:

Number of Papers published in non peer-reviewed journals:

(c) Presentations

Poster: Error rates for the identification of 9mm firearms using the IBIS, KB Morris, R Jefferys, E Dearth, E Fabyanic, JH Davis, 66th AAFS Meeting, Seattle WA, February 2014
Paper: The relevance of firing pin and breechface scores in the interpretation of firearms evidence, KB Morris, CE Dearth, and RL Jefferys, 66th AAFS Meeting, Seattle WA, February 2014
Poster: Creating a Bayesian network using normalized IBIS scores of .40S&W cartridge cases, Keith Morris*, Elizabeth Dearth, Eric Law, Roger Jefferys, & Emily Fabyanic, CBD-IAI, Williamsburg, VA, Fall 2014.
Poster: Glock model 21 consecutively fired cartridge cases shot variation, Keith Morris* & Catherine Hefner, CBD-IAI, Williamsburg, VA, Fall 2014.
Paper: The statistical analysis of firearms, Ultra Electronics Forensic Technology Inc., Montreal, Canada, November 2014.
Poster: Creating a Bayesian Network Using Normalized IBIS scores of .357 Magnum and .38 Special Cartridge Cases, KB Morris, E Law, R Jefferys, & E Fabyanic, 67th AAFS Meeting, Orlando, FL, February 2015
Poster: Using likelihood ratios for source attribution of Glock™ model 21 fired cartridge cases, C Hefner, & KB Morris, 67th AAFS Meeting, Orlando, FL, February 2015.

Number of Presentations: 8.00

Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received	Paper

TOTAL:

Number of Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received		Paper
08/20/2014	1.00	Keith Morris, Elizabeth Dearth, Roger Jeffrys. The Relevance of Firing Pin and Breechface Scores in the Interpretation of Firearms Evidence, American Academy of Forensic Sciences - 66th Annual Scientific Meeting. 20-FEB-14, . : ,
08/20/2014	2.00	Keith Morris, Elizabeth Dearth, Roger Jefferys, Joshua Davis, Emily Fabyanic. Error Rates for the Identification of 9mm Firearms Using the IBIS, American Academy of Forensic Sciences - 66th Annual Scientific Meeting. 20-FEB-14, . : ,
11/03/2015	3.00	Eric Law, Roger Jefferys, Emily Fabyanic, Keith Morris. Creating a Bayesian Network Using Normalized IBIS scores of .357 Magnum and .38 Special Cartridge Cases, AAFS 67th Annual Meeting. 18-FEB-15, . : ,
11/03/2015	4.00	Catherine Hefner, Keith Morris. Using Likelihood Ratios for Source Attribution of Glock® Model 21 FiredCartridge Cases, AAFS 67th Annual Meeting. 19-FEB-15, . : ,

TOTAL: **4**

Number of Peer-Reviewed Conference Proceeding publications (other than abstracts):

(d) Manuscripts

Received Paper

TOTAL:

Number of Manuscripts:

Books

Received Book

TOTAL:

Received Book Chapter

TOTAL:

Patents Submitted

Patents Awarded

Awards

Graduate Students

NAME	PERCENT_SUPPORTED	Discipline
Elizabeth Dearth	1.00	
Daksha Yadav	0.25	
Naman Kohli	0.25	
FTE Equivalent:	**1.50**	
Total Number:	**3**	

Names of Post Doctorates

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Names of Faculty Supported

NAME	PERCENT_SUPPORTED	National Academy Member
Keith Morris	0.25	
FTE Equivalent:	**0.25**	
Total Number:	**1**	

Names of Under Graduate students supported

NAME	PERCENT_SUPPORTED	Discipline
Roger Jefferys	0.50	Forensic & Investigative Science
Joshua Davis	0.10	Forensic & Investigative Science
Eric Law	0.50	Forensic & Investigative Science
Kylie Gordon	0.50	Forensic & Investigative Science
Emily Fabyanic	0.50	Forensic & Investigative Science
Shreya Kamath	0.25	Forensic & Investigative Science
Samantha Gissendaner	0.25	Forensic & Investigative Science
Heather Birks	0.25	Forensic & Investigative Science
Stephanie Martin	0.50	Forensic & Investigative Science
FTE Equivalent:	**3.35**	
Total Number:	**9**	

Student Metrics

This section only applies to graduating undergraduates supported by this agreement in this reporting period

The number of undergraduates funded by this agreement who graduated during this period: 8.00

The number of undergraduates funded by this agreement who graduated during this period with a degree in science, mathematics, engineering, or technology fields:...... 8.00

The number of undergraduates funded by your agreement who graduated during this period and will continue to pursue a graduate or Ph.D. degree in science, mathematics, engineering, or technology fields:...... 5.00

Number of graduating undergraduates who achieved a 3.5 GPA to 4.0 (4.0 max scale):...... 2.00

Number of graduating undergraduates funded by a DoD funded Center of Excellence grant for Education, Research and Engineering:...... 0.00

The number of undergraduates funded by your agreement who graduated during this period and intend to work for the Department of Defense 0.00

The number of undergraduates funded by your agreement who graduated during this period and will receive scholarships or fellowships for further studies in science, mathematics, engineering or technology fields:...... 2.00

Names of Personnel receiving masters degrees

NAME	
Roger Jefferys	
Eric Law	
Total Number:	**2**

Names of personnel receiving PHDs

NAME	
Total Number:	

Names of other research staff

NAME	PERCENT_SUPPORTED
Michael Bell	1.00
Theresa Joslin	1.00
FTE Equivalent:	**2.00**
Total Number:	**2**

Sub Contractors (DD882)

Inventions (DD882)

Scientific Progress

Introduction

In order to automate the process of matching a bullet with known firearms, Integrated Ballistics Identification System (IBIS) was developed. The IBIS system uses bullets and casings from case evidence from a crime scene and compares them to a database of known fired weapons. IBIS provides a relative score for each comparison, and a list of highest matching breechface scores as well as firing pin scores are generated as possible candidates for further comparison by firearm examiners. Beauchamp and Roberge generated a database of 500 pairs of cartridge cases (each pair fired from the same firearm) for each of the following calibers: 9mm, 32 auto, 45 auto, and 22. They computed a curve to predict the performance of the IBIS system as a function of the database size. They analyzed that the expected performance of IBIS decreases from 80% to 30-45% when the database size increases from 1000 to one million exhibits. They also concluded that the breechface, firing pin and ejector marks provide complementary information.

Traditionally in forensic science the terms class- and individual-characteristics are well used, but various interpretations and usages of these terms occur frequently in the literature. Generally classification is considered how results of the method are grouped. One can consider that the outcome of a method, especially in impression evidence, is a basic classification problem (match/non-match). This is in contrast to the class characteristics of the evidence.

The Association of Firearm and Toolmark Examiners (AFTE) has stated their “theory of identification” in three principles: (1) The theory of identification as it pertains to the comparison of toolmarks enables opinions of common origin to be made when the unique surface contours of two toolmarks are in “sufficient agreement”. (2) "Sufficient agreement” is related to the significant duplication of random toolmarks as evidenced by the correspondence of a pattern or combination of patterns of surface contours. Significance is determined by the comparative examination of two or more sets of surface contour patterns comprised of individual peaks, ridges, and furrows. Specifically, the relative height or depth, width, curvature and spatial relationship of the individual peaks, ridges, and furrows within one set of surface contours are defined and compared to the corresponding features in the second set of surface contours. Agreement is significant when it exceeds the best agreement demonstrated between toolmarks known to have been produced by different tools and is consistent with agreement demonstrated by toolmarks known to have been produced by the same tool. The statement that “sufficient agreement” exists between two toolmarks means that the agreement is of a quantity and quality that the likelihood another tool could have made the mark is so remote as to be considered a practical impossibility. (3) Currently the interpretation of individualization/identification is subjective in nature, founded on scientific principles and based on the examiner’s training and experience.”

According to this statement the concept of sufficient agreement is achieved when the agreement firstly exceeds the best agreement demonstrated between toolmarks known to have been produced by different tools and secondly the agreement is consistent with toolmarks known to have been produced by the same tool.

Best known non-matches and ROC curves

According to an online training program funded through NIJ by NFSTC and in collaboration with AFTE, the degree of correspondence which must be exceeded in order to reach sufficient agreement to effect an identification is the best known non-match (BKNM) as determined by each individual examiner and as produced by different tools. The individual examiner gains this experience during their initial training period rather than when they begin to perform their own casework examinations. Anecdotally, it is known that examiners do find better BKNMs during casework.

In order to understand the process, this concept will be elaborated upon. Given that each firearms examiner establishes their own BKNM, it seems plausible that since there are multiple examiners, there must be a range of BKNMs. This being true, the implication is that for a crime scene sample-known sample pair (CS-KN pair), we have two examiners, x and y, then each of these examiners will have their own BKNM threshold (BKNMx and BKNMy respectively). If we furthermore define the characteristics in congruence on the above-mentioned pair, the following threshold range of “sufficient agreement” can be defined as: BKNMx < Match between CS-KN pair < BKNMy

To further understand the implications of the AFTE theory of identification, another concept needs to be addressed. This concept may be illustrated by using consecutively matching striations method (CMS) as described by Biassoti in 1959. Since the introduction of this method, there has been much debate between the so-called pattern matchers and line counters.

According to Nichols, one of the pitfalls of the CMS method (versus the pattern matching method) is the large degree of false exclusions. The may be restated as the exclusion of a match based on CMS when a pattern matcher would call it a match. Is this an unexpected behavior? In order to understand the behavior, one needs to return to the discriminating power of a method.

The discriminating ability of a method can be described by its sensitivity and specificity. The sensitivity of a method is its ability to detect a condition when the condition is present (or calling a match, a match). Specificity is the ability of a method to detect an absence when the condition is not present (or calling a non-match a non-match). The sensitivity of a method is equivalent to the true positive rate (tpr), whilst the specificity is equivalent to the true negative rate (tnr). Given the focus on methods generally used in forensic science and that they are subject to the Daubert criteria, it is important to understand the known or potential rate of error in the method. Generally courts are interested in the false positive rate (fpr) and their false negative rate (fnr) of a method.

Assume the performance of a method for the comparison of cartridge cases as typically exercised in a forensic laboratory. The arbitrary measure on the x- axis represents the result of the comparison between many pairs of cartridge cases for which the ground truth is known. For each cut-off on the x-axis, a finite value for the fpr and fnr is given. The concept of a BKNM can now be defined in terms of the fpr. Let BKNMx = 1500 and BKNMy = 2000. Based on these assumptions one can consider three situations: (1) If the CS-KN pair has a score of 1000, both examiners will classify this as a non-match, (2) if the CS-KN pair has a score of 1250 then examiner x will classify it as a match whilst examiner y will classify it as a non-match, and (3) if the CS-KN pair has a score of 2500 both examiners will classify it as a match. Let the standard (perfect) BKNM be defined as BKNMs. Thus, BKNMi ¿ BKNMs, fpr¿ 0. Thus, irrespective of a CMS or pattern matching approach, the concept of eliminating false exclusions is unavoidable. Conversely, if the number of false exclusions you reject increases the fpr rate will increase.

What is a “match”?
To ensure clarity in further discussion of this project it may be useful to provide some working definitions of common terms. The term “match”, in forensic science, generally means that some item of evidence is attributed to a particular source. This process generally implies that the true state is unknown. In firearms examination, in particular, match generally implies a same gun source attribution whereas a non-match implies an attribution of different sources for the cartridge case and firearm (different gun). Matches between objects of the same class are usually achievable when the within class variability (intra-variability) is significantly smaller that the between class variability (inter-variability).
If m cartridges are fired from firearm M and n cartridges from firearm N, then cartridge mi matches mj (as does ni and nj because of their common source) but cartridge mi does not match cartridge ni. This is true irrespective of how many features an examiner may or may not find. The match status in examinations is based on features.

Inter- and Intra-variability
Comparison and identification is thus ultimately dependent upon the intra-variability and inter-variability of cartridge cases from various firearms. Two new .45 ACP caliber pistols (Glock 21 Gen4 and Taurus 24/7 G2) were used to fire 50 cartridges of the same brand (Federal American Eagle). Each cartridge was marked with a permanent marker to identify it as the nth shot fired through the firearm. This seemingly simple inter-cartridge compassion becomes quite complex. It appears as if there is no change in the scores even though there is a large variation in the confidence intervals between shots. When multiple shots from the same firearms are entered in the IBIS system, one would expect that the cartridges previously entered from the same firearm would feature high in the generated candidate list. Thus with each additional entry, p, then p-1 candidates would be expected in the candidate list for the particular firearm.

Ideally at a separation of one, one would expect 49 values, at separation two, one would expect 48 values, and so on. The dotted lines represent the 95% confidence interval for each separation accounting for sample size. Evaluating these results for the Glock .45 ACP caliber pistol would seem to suggest that there is no change between the various separations although the distributions within each separation are relatively large.

As would be expected, there is a general decrease in score with an increase in rank. At high rank (low numerical values), the match scores have a higher minimum than the non-matching scores. The match scores have a higher density a high rank. The reverse is true for the non-match scores. In general the firing pin non-match scores seem high, indicating lower discrimination.

Sample size
Firearm examiners will usually test-fire a suspect firearm two to five times using ammunition similar to that found at the crime scene. The actual number fired is determined by laboratory policy and the experience of the firearms examiner. The examiner will then select a cartridge case from the set which is deemed to be representative of the suspect firearm. This cartridge case is then used as the known in the comparison process performed on a comparison microscope. Cartridge cases are generally entered into the IBIS system given two scenarios: (1) After a comparison is made, and the suspect cartridge case is deemed not to have been fired by the suspect firearm, or (2) after a single or group of cartridge cases is examined, the firearms examiner may select one or more cartridge cases for submission to IBIS.
In this project a sample of 100 cartridge cases, in most cases, were used to evaluate the variability of same gun (Hd) and different gun (Hp) scores. In order to make use of a sample of cartridge cases from a firearm to develop a same gun distribution, the sample distribution must be representative of the actual distribution of same gun scores.

The actual and sample distributions were compared and the variability between the two was computed using the sum of the squares. Various sample sizes were used to assess the effect of the sample size in approximating the actual distribution.

The density distributions of the same gun and different gun scores were simulated for pistol X45399. The simulations also assessed the influence of sample size. It must be noted that 10 cartridge cases will result in 45 pairwise comparisons, and thus 45 breechface scores, to define the same gun distribution.

Performance of 9mm firearms
Approximately 100 cartridge cases fired by the 9mm firearms (35 pistols, 2 carbines, and 1 revolver) were submitted to IBIS and

the resulting breechface and firing pin scores were analyzed using R and RStudio. The data were divided by model of firearm and a receiver operating characteristic (ROC) curve was computed. The area under the ROC curve was also computed. All ROC calculations were performed using the ROCR package in R.

All of the areas under the curve (AUC) for the receiver operating characteristic (ROC) curve were provided. These data are for the breechface scores (BF), the firing pin scores (FP) and for their product (FP*BF). The most discriminating measures for a particular firearm were assessed. An AUC of .500 indicates that the method of classification is equal to a coin toss. A method with an AUC of 1 indicates a method which has perfect classification performance. The error rate curve illustrates how the particular cutoff (on the x-axis), in this case (BF score, FP score, and their product) affects the false positive rate (fpr) and the false negative rate (fnr). The point at which they cross is known as the equal error rate (EER). Forensic scientists would like to have a low fpr and thus world generally work to the right of this position with some tradeoff for the fnr.

IBIS scores for the SCCY CPX II pistols performed very badly as a classifier for the same gun/different gun scenario. The best performers were for the Ruger SR9, SigSauer P250, and the Taurus Millennium Pro 111. Of the 18 models tested, four had breechface score as the best performer, four with the firing pin score, and 10 with the product of both. This is illustrative than in many case both should be considered. This is not easily achieved with the current configuration of the IBIS. Overall, the product of the scores is the best performer in classification.

Data analysis of IBIS breechface and firing pin scores

In order to define a methodology for the analysis of the data the following standard definition (where available) were used to develop a useable definition. According to the EURACHEM Guide, repeatability may be defined as: “Precision under repeatability conditions, i.e. conditions where independent test results are obtained with the same method on identical test items in the same laboratory by the same operator using the same equipment within short intervals of time”. Furthermore, it states that the repeatability standard deviation is the “standard deviation of test results obtained under repeatability conditions”. In addition, it notes that the repeatability standard deviation is a “measure of dispersion of the distribution of test results under repeatability conditions. Similarly ‘repeatability variance’ and ‘repeatability coefficient of variation’ could be defined and used as measures of the dispersion of test results under repeatability conditions.” For this study the definition of repeatability is utilized is a variation on the definition as proposed by Eurachem. The measure of precision used to evaluate the repeatability is the coefficient of variation for both breechface and firing pin scores.

Precision under repeatability conditions, i.e. conditions where independent test results are obtained with the same acquisition method on repetitive imaging of an identical test item in the same laboratory by the same operator using the same equipment within short intervals of time.

Also from the EURACHEM Guide, Reproducibility may be defined as: “Precision under reproducibility conditions, i.e. conditions where test results are obtained with the same method on identical test items in different laboratories with different operators using different equipment.” It also notes that a “valid statement of reproducibility requires specification of the conditions changed. Reproducibility may be expressed quantitatively in terms of the dispersion of the results”. The reproducibility standard deviation is defined as the “standard deviation of test results obtained under reproducibility conditions”. Similar to repeatability, this is a “measure of dispersion of the distribution of test results under reproducibility conditions. Similarly ‘reproducibility variance’ and ‘reproducibility coefficient of variation’ could be defined and used as measures of the dispersion of test results under reproducibility conditions.” For this study the definition of reproducibility is utilized is a variation on the definition as proposed by Eurachem. The measure of precision used to evaluate the reproducibility is the coefficient of variation for both breechface and firing pin scores. Precision under reproducibility conditions, i.e. conditions where test results are obtained with the same acquisition method on separate imaging of an identical test item in the same laboratory with the same operators using the same equipment.”

Repeatability

The data were sliced to obtain select the breechface (BF) and firing pin scores (FP) of each analyst against themselves as per the adopted definition of repeatability. Since the coefficient of variation (CoV) takes both the mean and standard deviation into account, this implies that an analyst with a low mean and a low standard deviation could have the same CoV as an analyst with a higher mean and a higher standard deviation. The max CoV BF is less than 11%, whilst the max CoV for FP is less than 30%. This variability between examiners may seem high for the FP CoV, but it must be remembered that the score values obtained in this study are extremely high scores, not usually seen in casework. It may also amplify the fact that small changes in light may have a considerable impact on the net score.

Reproducibility

Similar to the repeatability study, the data were sliced to obtain select the breechface (BF) and firing pin scores (FP) of each analyst against themselves as per the adopted definition of reproducibility. The max Cov BF is slightly less than 12% (similar to the repeatability value), whilst the max CoV for FP is less than 30%. Apart from one examiner (NMC, a new student), the rest of the CoV’s are more clustered.

Blind studies

For each of the 5 blind samples, 4 sets of data (lowest numerical rank, highest BF, highest FP, and highest FP*BF) were used to select the data. Once the data was sorted a full set of Rank, FP and BF were entered into the BN. Results which are to the right of the y-axis support the selection of a particular model. Results to the left of the y-axis provide support for the particular model not being the one which fired the question cartridge case. The extent to which the results deviate from the y-axis demonstrate the magnitude of agreement with the proposition. It can be seen that there is very strong evidence to support the proposition that the cartridge case was not fired from the Ruger SR9, the HiPoint 995TS, or the Taurus Millennium Pro 111. There is moderate evidence to support the proposition that the cartridge case was fired from a Smith & Wesson SD9-VE. There is limited evidence to support the proposition that the cartridge case was fired from a Keltec PF9 and that it was not fired by a Keltec P11. Given these results, the ability to infer the make and/or model of a firearm from IBIS scores seems limited at present.

Comparison
The comparison data were evaluated in a few ways. A traditional statistical approach and a Bayesian approach were undertaken. For example, the match (matches between K1, K2, and K3), non-match (K1, K2, K3, and Q1 versus a pre-existing case in the database), and unknown (Q1 versus K1, K2, and K3) density distributions are given for both the FP and BF scores. By inspection of the FP distributions it appears that the unknown distribution is similar to the of the non-match distribution From the Bayesian Network perspective and given the selection of the data sets to compare, the LR for each instance was computed and is given below:

Set	Rank	BF	FP	LR	Verbal	Ground Truth	Known	Questioned
1	306	60	67	5.3	limited evidence to support Hp	Same Gun	Ruger P95DC	Ruger P95DC
1	760	63	49	2.6	limited evidence to support Hp	Same Gun	Ruger P95DC	Ruger P95DC
1	172	38	71	0.6	limited evidence to support Hd	Same Gun	Ruger P95DC	Ruger P95DC
2	48	24	41	1.1	limited evidence to support Hp	Different Gun	HiPoint C9	HiPoint 995
2	95	25	37	0.4	limited evidence to support Hd	Different Gun	HiPoint C9	HiPoint 995
2	27	10	48	1.4	limited evidence to support Hp	Different Gun	HiPoint C9	HiPoint 995
3	698	28	32	0.2	limited evidence to support Hd	Different Gun	Springfield XD9	HiPoint C9
3	875	31	23	0.2	limited evidence to support Hd	Different Gun	Springfield XD9	HiPoint C9
3	586	21	35	0.2	limited evidence to support Hd	Different Gun	Springfield XD9	HiPoint C9

Analysis of NIST standard cartridge cases
In order to assess the standard performance of IBIS, five standard cartridge cases from NIST were used. Each standard reference material (SRM) was entered into IBIS 10 times by each of three users (EBF, RLJ, and EFL). Each of these users has more than 12 months experience entering cartridge cases into IBIS. The SRM's used were 2P2333, 2P2335, 2P2415, 2P4316, and 2P6325. These were run as normal 9 mm Luger cartridge cases and the candidate lists were processed in the usual manner. Recovery values < 100% are coded in pink. For the 150 samples submitted), there are 11,175 possible comparisons 150C2. Since all of the samples are re-correlated, there are twice the number of comparisons (a vs. b, and b vs. a) giving 22,350 comparisons. Of these, 63 comparisons were not recovered by the IBIS system. It is interesting that the recoveries are asymmetrical. For example, EBF-2P2415 vs. RLJ-2P2333 has a recovery of 98%, whilst RLJ-2P2333 vs. EBF-2P2415 has a recovery of 100%. It is clear that self-recovery will not occur (a. vs. a.), thus the diagonal has no instances of comparisons returned. In all cases, there will not be a recovery for the sample against itself. In most instances, the candidate lists yield at least 2,000 candidates. All of the lists contained non-match data. It is unclear the recovery loss, although small (~0.28%), occurs. In four, of the nine comparisons perfect recovery was achieved. Of the remaining five, the average recover was 98.30%. Interestingly, no analyst achieved a 100% recovery against their own submissions.

The firing pin (FP) and breechface (BF) scores, as well as their product (FPBF) were evaluated according to their receiver operating characteristic (ROC) curves. For these data both the BF and FPBF have perfect discrimination (AUC = 1.0), whilst the FP is near perfect. It can be seen that if an SRM has a BF score more than about 25 then it is a Match, without the influence of false negatives at higher scores. In Figure 84, it can be seen that if an SRM has a BF score more than about 65 then it is a Match. If the score increases to about 80, then the possibility of false negatives becomes real.

Normalization study
During a meeting with the representatives of Ultra Electronics Forensic Technology Inc. (FTI) the concept of score normalization was discussed. It was also stated that the ranks of the breechface and firing pin scores are more discriminating. Up to this point, only the rank of the firing pin has been used in calculations. According to FTI, the IBIS correlation process is broken down into two sub-processes, coarse and fine correlation. The course correlation is a fast but less accurate correlation. The objective of this process is to reject rapidly the matching candidates. This process is performed independently on the breechface and firing pin scores. The top 10% of candidates from each list all then processed using the fine correlation procedure. The scores calculated during this fine correlation process all the scores which are provided by the system. The scores calculated during the course correlation are not used further in the process. This approach can result in a candidate having a high breechface score and a low firing pin score for example. In this case, the candidate was identified through the course correlation of the breechface scores.

It is seen that both the firing pin score and the breechface score perform equally well as classifiers. There is a concentration of matching scores (pink dots) with high ranks (low values) for both firing pin and breechface. There are also bands across the

axes at high ranks for each, but low ranks for the other. In the bulk of the data there are both match and non-match data at relatively low ranks.
The following raw and derived metrics for breechface and firing pin scores used to evaluate method efficacy: auc.BF, auc.FP, auc.BFFP, auc.FP_Rank, auc.BF_Rank, auc.BFFP_Rank, auc.BFFP_Score_Over_BFFP_Rank, auc.BF_norm, auc.FP_norm, and auc.BFFP_norm. These metrics were used to assess their applicability as a classifier of the IBIS data. In general, the scores can be categorized into three categories namely, firing pin related scores, breechface related scores, and combinations of firing pin and breechface scores. The rank scores are associated with the particular metric since ranks are simply the rank order of those scores.

There are indications of the separation and overlap between same source and different source guns. All different source guns are of the same caliber and of multiple makes and models (to include examples of the same make and model). When one considers the non-match distribution it can be seen that the firing pin scores reach a maximum value of just under 100, whilst the breechface scores reach a maximum at approximately 125. There is significant overlap of match and non-match scores in this region with a strong cluster at very low scores, and a high-density cluster centered around (25, 25). As one moves up the diagonal when reaching the (50, 50) position, the matches seem to separate from the non-matches. However, the density of matches in this region appears significantly lower.

2D versus 3D study
The data was separated by firearm in order to analyze the intra- and inter-variability between the same makes as well as the same models with different identifiers (serial numbers). The SCCY CPX II firearms performed the best with regards to 2D BF scores; however, they did not perform the same and have two separate maximums and minimums. This observation indicates that BF is has the best discriminatory power for SCCY CPX II firearms. The Springfield XD9 firearms performed highest with regards to FP scores and lowest with BF scores, indicating that FP has the better discriminatory power. Similar to that of the SCCYs, these two firearms of same make and model did not perform the same. There were three Keltec firearms analyzed of three different models: P11, Sub-2000, and PF9. All three performed the best with respect to FP scores and the worst with BF scores indicating a class characteristic that the FP has a higher discriminatory power than the BF. The Sub-2000 and the PF9 performed similarly both having auc.FP_3D as the highest score and auc.BF_norm as the lowest score, whereas the P11 had the highest value with auc.RankFP and the lowest with auc.BF. The two Ruger firearms, LC9 and SR9, performed similarly in the fashion that the FP had the highest scores and the BF had the lowest. The LC9 performed the same across five categories of FP scores resulting in a value of 1. The only HiPoint performed best using the 2D Sidelight feature of BF analysis and the worst at the standard BF position. Unlike the other makes, it is unclear if BF or FP is a more discriminatory feature of a cartridge case from a HiPoint firearm. The Arcus D98 and the Taurus 24/7 G2 can be better identified from the FP impression than from the BF, which is reflected in their minimum and maximum scores. Overall, with respect to all the firearms examined, every minimum value is derived from the BF scores (2D, 3D, or normalized).

For both systems, the Ruger LC9 (X43521) had the lowest value resulting from the BF scores (0.447 WVU and 0.506 for FTI). It is interesting to note that the Ruger LC9 was the worst performance in both 2D BF categories while the Ruger SR9 performed the highest.

It can be seen that the BF of a Ruger SR9 has a high discriminatory power. The quality of performance of the BF impressions is not the same across different models of Ruger firearms. If the analysis of the SR9 had not been included in this study, one might assume that poor performance of BF scores is a class characteristic of all 9mm Ruger firearms.

A general comparison of performance of the two systems was underetaken. The linear regression and the y=x indicates the similarity in scores. The variability is assigned to user and sample orientation. It appears as if the breechface match scores follow the y=x line and the regression is weighted to the non-match scores. The firing pin match scores follow the regression line, but at higher scores the FT system attributes higher scores than the WVU system.

USACIL test set
Likelihood ratios were calculated for the test and evidence samples. Tests 1, 3, and 7 are similar in that they contain ranks, whilst tests 2 and 4 do not. Each of the tests is conditioned on the firing pin type of the submitted sample. The states of the Firing_Pin_Type_Sample node are Circular and Glock. The conditioning is necessary to obtain the correct prior odds. It is furthermore necessary that the node Firing_Pin_Type_DB will be conditioned similarly. This assumes that all of the firing pin types in the IBIS are correctly entered. Searching of the IBIS database is conditioned on Firing Pin Type in the system. In other words, when a sample is entered into IBIS, its firing pin type will result in the search be launched against cartridge cases with the same firing pin type. In some instances, the background data were incorrectly classified regarding their firing pin type. Unfortunately, the IBIS does not use the presence or absence of a drag mark as a classifier with in its database. Where possible, the presence or absence of a drag mark in the background data has been entered into the test sets. In instances where the state of the drag mark is unknown, the node will be given a state of Unknown. The conditioning on drag marks will separate firearms with a blowback action (e.g. HiPoint C9) from those with a recoil lock system (e.g. Ruger SR9).
Figure 121: Comparison of the log likelihood ratios (Test1 and Test2) of cartridge case U01
In Figure 121, the logarithm of the likelihood ratios (LLR) for Test 2 are plotted against the LLRs for Test 1. The conditioning factors for these results are given in Table 43, and the classification of the evidence cartridge case is given in Table 44. Both

panels of Figure 121 present the data separated by the value of the Model_DB node. The upper panel (and all subsequent similar figures) provides the LLRs with the Drag_Mark_DB node having a state of Yes, whilst the lower panel provides the data for the Drag_Mark_DB node having a state of No. From Table 44 it is known that the evidence cartridge case, U01, does not feature a drag mark. Thus the plots given in the upper panel represent nonmatching candidates, whilst those in the lower panel represent potentially matching candidates. In this case, the lower panel will also include sections labeled “Test” and ”Evidence”. The “Test” section provide the LLRs for the test versus test samples, whilst the “Evidence” section provide the LLRs for the ‘test versus evidence’ and ‘evidence versus test’ samples. For these results, it must be borne in mind that the results are not conditioned on the Drag_Mark_Sample node.

USACIL test set revisited

Sample Set	Known Firearm Make/Model
U01	Sig Sauer P228
U02	Sig Sauer P226
U03	Sig Sauer P226
U04	Glock 19
U05	Ruger P89DC
U06	Ruger P89DC
U07	Glock 19
U08	Smith &Wesson SW9VE
U09	Smith &Wesson SW9VE
U10	Taurus PT 24/7 PRO
U11	Taurus PT 24/7 PRO
U12	Taurus PT 709
U13	Springfield Armory XDM-9

After receipt of the information, a reassessment of the data provided resulted in the adaption of the Bayesian network to differentiate between the presence of a drag mark on the prime of a cartridge case and the type of action of the firearm. Generally, two main types of pistol actions are encountered within the data set. Blowback action is a type of design in which the there is no locking of the bolt. The breech is held closed only by the weight and inertia of the bolt, with some slight assistance from the recoil spring, until the bullet leaves the muzzle. In a recoil action (locked breech) pistol, the barrel and slide are securely locked together at the moment of firing. They travel backward together until the barrel unlocks, forced down by a link or inclined plane, and continues rearward under its own momentum. A HiPoint C9 pistol has a blowback action, whilst a Ruger SR9 has a recoil action. Drag mark are generally only found on cartridges fired by a recoil action pistol. Some recoil action pistols seldom generate a drag mark on their cartridge cases e.g. SIgSauer P250.

A SCCY CPX II is selected as the model of the firearm. This pistol has a recoil action and thus has a locked breech. The Yes state of the node ActionLB_Sample becomes 100%. When the Match node is instantiated to Yes, the ActionLB_DB updates to Yes =100%. A match can only be between the same SCCY CPX II pistol, which are a locked breech action. For the nodes Drag_Mark_Sample = Yes (42.4%) and Drag_Mark_Sample = No (57.6%) indicating that the presence of drag marks on these samples is not well replicated. The inference to be made is that if a fired cartridge case was found from the SCCY CPX II pistol there is a 42.4% probability that it will have a drag mark.

Baldwin test set

In a study conducted by Baldwin et al. 25 Ruger SR9 pistols were conditioned by firing 200 cartridges in each pistol. Thereafter 800 cartridges were fired through each pistol and collected. Sets of one “questioned” cartridge case and three “known” cartridge cases were set up by the Baldwin group and sent out to firearms examiners for further analysis. Twenty sets were selected by the Defense Forensic and Biometrics Agency (DFBA) and submitted for analysis.

Maximum LLRs for all Baldwin data

Max LLR

Sample	Number of Records	LLR Test 1	LLR Test 2	LLR Test 3	LLR Test 4	LLR Test 7	Max LLR	Verbal Scale Value
Set 01	6	2.00	-0.09	1.37	0.08	1.61	2.00	Evidence strongly supports Hp
Set 02	1	-0.69	-1.28	-1.41	-1.24	-1.65	-0.69	Evidence weakly supports Hd
Set 03	6	3.58	2.05	2.74	2.18	2.86	3.58	Evidence very strongly supports Hp
Set 04	2	-0.97	-1.38	-1.69	-1.34	-1.88	-0.97	Evidence weakly supports Hd
Set 05	1	0.33	-0.95	-0.41	-0.93	0.50	0.50	Evidence weakly supports Hp
Set 06	4	2.01	-0.45	1.46	-0.37	2.29	2.29	Evidence strongly supports Hp
Set 07	4	1.42	-0.46	0.72	-0.41	1.66	1.66	Evidence supports Hp
Set 08	4	2.15	0.15	1.57	0.38	2.22	2.22	Evidence strongly supports Hp
Set 09	0							
Set 10	3	0.66	0.10	-0.07	0.13	1.29	1.29	Evidence supports Hp
Set 11	0							

Set 12	9	2.37	0.40	1.89	0.70	1.96	2.37	Evidence strongly supports Hp
Set 13	5	2.20	0.29	1.63	0.51	1.53	2.20	Evidence strongly supports Hp
Set 14	4	-0.15	-1.28	-0.87	-1.24	-0.52	-0.15	Evidence weakly supports Hd
Set 15	6	1.33	-0.84	0.60	-0.79	1.61	1.61	Evidence supports Hp
Set 16	6	3.67	1.64	2.90	1.62	3.07	3.67	Evidence very strongly supports Hp
Set 17	0							
Set 18	3	0.74	-0.69	0.02	-0.62	1.26	1.26	Evidence supports Hp
Set 19	6	2.73	0.78	2.23	1.09	2.69	2.73	Evidence strongly supports Hp
Set 20	2	1.90	-0.21	1.27	-0.03	1.15	1.90	Evidence supports Hp

Results of the determination of the log likelihood ratios (LLR) for the evidence vs test samples in each of the sets were provided. The number of records returned indicates the test/evidence comparisons which were returned by IBIS. For Set 09, Set 11, and Set 17 no records were retuned. In these data, all of the records from the Ruger SR9 study previously entered into IBIS were removed from the candidate lists and the firing pin and breechface ranks were recalculated without those data. In the plots the Model DB of unknown contains all comparison data between sets. For this analysis no prior information regarding the test firearms has been considered (i.e. the make and model of the gun is unknown).

For all of the SR9s, firearm X96651 has a large number of results since it was used in three of the twenty tests. In the Unknown firearms, four results have high LLR values. This results in comparisons between two of the question cartridges belonging to elimination sets (SET05-Q1: SET12-K2, SET12-K3 (X96385) and SET11-Q1: SET18-K2, SET18-K3).

Table 46: Association of likelihood ratio with verbal equivalent (Evett & Buckleton)

LLR of Evidence C	Conclusion
LLR = 0	The evidence is neutral
0 < LLR <= 1	The evidence slightly supports C
1 < LLR <= 2	The evidence supports C
2 < LLR <= 3	The evidence strongly supports C
3 < LLR	The evidence very strongly supports C

The verbal scales for LLR's are applied and compared to the Truth and Baldwin results. These data are given as follows: each question sample per set is associated with the LLR's of each test and each known cartridge case responding to a search on IBIS. The columns entitled "Evidence…" are the verbal scales associated with the LLR in the preceding column. These should be read as "The evidence _____ supports sgp/dgp". The "same gun proposition" (spg) and "different gun proposition" (dpg) are abbreviated for brevity. The cells highlighted in light green indicate that the LLR is in support of the Truth-value. Those in pink do not support the Truth-value. The empty cells provide strong support either for or against the Truth-value as per their color. It should be noted that some questioned samples have LLR's both in support and against the Truth-value indicating the variability

Figure 240: LLR results and Verbal scales

Once the data for the conditioning study are added back into the data set, there are more test results by type. These results are given in Figure 241. These results clearly indicate the improvement of the LLR's for the matching data. This underlines that the variation in the markings are better represented through an increased sample size when IBIS is used as the measuring instrument.

Set	Letter	Serial Number	Truth	Baldwin Results	LLR(Test 7) Ability
Set 01	D3	X96664	Same Gun	Inconclusive	˜
Set 02	D5	X96667	Same Gun	False Negative	correct
Set 03	A1	X96383	Same Gun	Inconclusive	correct
Set 04	B5	X96592	Same Gun	Inconclusive	˜
Set 05	D2	X96663	Different Gun	False Positive	correct (X96385)
Set 06	B5	X96593	Same Gun	Inconclusive	˜
Set 07	E3	X96689	Same Gun	Inconclusive	correct
Set 08	C5	X96651	Same Gun	Inconclusive	correct
Set 09	C1	X96594	Different Gun	False Positive	correct (X96719)
Set 10	C3	X96620	Same Gun	Inconclusive	˜
Set 11	B5	X96593	Different Gun	False Positive	? (Firearm not in DB)
Set 12	A2	X96385	Same Gun	Inconclusive	correct
Set 13	C5	X96651	Same Gun	Inconclusive	correct
Set 14	C3	X96620	Different Gun	False Positive	correct (X96669)
Set 15	E2	X96681	Different Gun	False Positive	correct (X96590)
Set 16	C5	X96651	Same Gun	False Negative	correct
Set 17	E5	X96719	Different Gun	False Positive	correct (X96593)
Set 18	D4	X96665	Different Gun	False Positive	correct (X96383)

Set 19	E4	X96718	Same Gun	Inconclusive	correct
Set 20	E2	X96681	Same Gun	False Negative	correct

Bayesian network website
WVU has conducted extensive research and data analysis on various firearms, including cartridge case comparisons. One of the best ways to describe data is by fitting it to a statistical model. Bayesian statistics offers an approach with a natural framework to deal with parameter and model uncertainty. The end goal of Bayesian analysis is to provide a distribution for the knowledge gained (i.e. what was learned) about the parameter from the data. Netica, a Norsys Software Corp program, is a simple, reliable, and high performing Bayesian network development software. A Bayesian network is a model that reflects the states of the given population being modeled and describes how those states are related by probabilities. The aim of this chapter is to provide an easy to follow user manual for setting up and utilizing the Netica-based cartridge case individualization web interface.
Case 1 utilizes the breechface (BF), firing pin (FP), and ranks from the IBIS system scores to find the match probability and likelihood ratio values.

The goal of case 2 is to predict the best possible match of the make and model of an unknown firearm. This situation could be applicable when there is no firearm recovered, from the scene or persons of interest, but a cartridge case has been collected.

The goal of case 3 is to determine the likelihood ratio of a known firearm.

Technology Transfer

Interpretation of cartridge case evidence using IBIS and Bayesian networks

Keith B. Morris, Eric F. Law, Roger L. Jefferys, and Elizabeth C. Dearth

Department of Forensic & Investigative Science
West Virginia University

Table of contents

Sponsors and Disclaimers

Research was sponsored by the U.S. Army Research Office and U.S. Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-12-2-0056.

The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office, Army Research Laboratory, the Defense Biometrics and Forensic Agency, the Defense Forensic Science Center, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

This project is funded by the Assistant Secretary of Defense for Research and Engineering ((ASD(R&E)) and the project is managed by the Defense Forensics and Biometrics Agency, Defense Forensic Science Center, Office of the Chief Scientist.

Acknowledgements

The study would have been impossible without the willing participation of many individuals. The study was managed by the Defense Forensics and Biometrics Agency, Defense Forensic Science Center, Office of the Chief Scientist. A sincere word of thanks to Rick Tontarski (Chief Scientist), and program mangers Karen Olson, Henry Maynard, and Garold Warner and other members of the OCS staff located at the Defense Forensic Science Center who helped to shape the project and provided much guidance. Dr. Michael "Jeff" Salyards, Executive Director, DFSC, is thanked for his help and motivation during the early stages of this project.

I would like to thank all of the staff, graduate and, undergraduate students who helped with this project, specifically Mike Bell, Roger Jefferys, Eric Law, Elizabeth Dearth, Shreya Kamath, Emily Fabyanic, Samantha Gissendaner, Stephanie Martin, Kylie Gordon, Heather Birks, Naman Kohli, Daksha Yadav, Theresa Joslin, and Joshua Davis.

Abstract

This project was focused on interpreting IBIS data to provide for a statistical analysis of firearm and toolmarks. The IBIS system provides an objective measure of the correlation of images of two breech face or firing pin impressions on two cartridge cases. This may be restated that better correlations are represented by higher scores. Cartridges fired by the same gun should thus result in similar images and thus higher scores. Cartridges fired by the different guns should thus result in dissimilar images and thus lower scores. The generated scores were transformed, along with characteristic information regarding the firearm and related information, into a Bayesian network. A Bayesian network is a directed acyclic graph, which when constructed from a forensic perspective, will allow one to assess the value of evidence based upon two propositions, viz. whether the cartridge was fired by the same gun or by another gun (different gun). The relative value of these questions can be transformed into what is known as a likelihood ratio. This allows a forensic scientist to provide insight to courts and investigators as to the value of the evidence.

This study indicated that a better understanding is required for the causes of the relatively high variability in cartridges cases fired by the same firearm as measured by IBIS scores. An initial attempt to answer this question was done by simulating the minimum number of cartridge cases required to produce a distribution equivalent to that of the firearm. The breech face (BF) and firing pin (FP) scores, as well as their product (BFxFP), generated by the IBIS were used to assess the ability of the system to classify an "unknown" cartridge case into a *same-gun* or *different-gun* category. The IBIS system does not provide for an easy means to use the combination of the BF and FP scores. The ability to order candidate lists through the combination of scores will be of value to firearms examiners (especially so in the 3D system). Generally, all of the classifiers performed well but the SCCY CPX II pistols were the worst in all three measures. This was due to markings which were difficult for IBIS to interpret, but would be easy identifications for a firearms examiner. The reliability of the IBIS system was assessed using the NIST Standard Reference Material® 2461 (standard cartridge case). A 2D IBIS heritage system was compared to the new 3D IBIS system and found that the results were very well correlated. Twenty sets of known and questioned cartridge cases, from a large collection which had been analyzed by operational firearms examiners, were examined and tested using the Bayesian networks. Out of the 20 comparisons, there were eight true positives, seven true negatives, five false negatives, and zero false positives. In all instances of eliminations, the support for the different-gun hypothesis was, at minimum, strong.

Overall, this study supports the interpretation of IBIS results through Bayesian networks. Improvements to the manner in which results are made available to the user will allow for more in-depth analysis of such results.

Introduction

It is estimated that the total civilian population of the world has over 650 million firearms out which the people of the United States alone possess around 270 million firearms[1]. According to the FBI, firearms were used in a staggering 69.3% of the total reported homicides in 2012 and 41% of the total robbery cases in 2012[2]. Therefore, the identification of firearms used in these cases to apprehend the suspect is imperative. It has been found that cartridges discharged from a firearm leave marks on the bullets and cartridge cases which are often collected as evidence from the crime scene. The most prominent marks are usually left on the soft primer of the fired cartridge case. These impressions, identified as breech face and firing pin marks can be unique to a firearm and can cause the cartridge case to be identified to a particular firearm with high degree of certainty. In order to determine if trained firearm examiners are able to link fired bullets to their firearms, Hamby et al. evaluated 507 firearms examiners from over 20 countries[3]. They were asked to compare unknown fired bullets to the rifled barrels and out of 7,605 unknown fired bullets, 7,597 were correctly matched to the known bullets. Their study concluded that there are identifiable features on the bullets that allow their identification from the gun that fired it.

In order to automate the process of comparing bullets with known firearms, Integrated Ballistics Identification System (IBIS) was developed[4]. The IBIS system uses bullets and casings from case evidence from a crime scene and compares them to a database of known fired weapons. It is also possible to compare bullets and cartridge cases from a crime scene to those from other scenes. IBIS provides a relative score for each comparison, and a list of highest matching[5] breech face scores as well as firing pin scores is generated as possible candidates for further comparison by firearm examiners. Beauchamp and Roberge generated a database of 500 pairs of cartridge cases (each pair fired from the same firearm) for each of the following calibers: 9mm, 32 auto, 45 auto, and 22[6]. They computed a curve to predict the performance of the IBIS system as a function of the database size. They analyzed that the expected performance of IBIS

---

[1] Karp, Aaron. Estimating Civilian Owned Firearms. September 2011. http://www.smallarmssurvey.org/fileadmin/docs/H-Research_Notes/SAS-Research-Note-9.pdf.

[2] FBI. 2013. http://www.fbi.gov/about-us/cjis/ucr/crime-in-the-u.s/2012/crime-in-the-u.s.-2012/offenses- known-to-law-enforcement/expanded-offense/expandedoffensemain.

[3] James E. Hamby, David J., James W. Thorpe. "The Identification of Bullets Fired from 10 Consecutively Rifled 9mm Ruger Pistol Barrels: A Research Project Involving 507 Participants from 20 Countries." AFTE Journal, 2009: 99-110.

[4] http://www.forensictechnology.com/ibistrax

[5] From the IBIS system perspective.

[6] Beauchamp, Alain, and Danny Roberge. "Model of the Behavior of the IBIS Correlation Scores in a Large Database of Cartridge Cases." 2005.

decreases from 80% to 30-45% when the database size increases from 1000 to one million exhibits. They also concluded that the breech face, firing pin, and ejector marks provide complementary information.

Researchers have tried to develop various systems that have better imaging and comparison methods than IBIS. In their study, Cork et al. analyzed and summarized that a national reference ballistic imaging database is not currently possible to build, as the existing imaging methods do not have sufficient discriminatory power to identify firearms on a large scale[7].

Recently, Petraco et al. focused on striation patterns from tools as well as cartridge cases from firearms[8]. They used 37 different Glock pistols to collect data for a total of 186 cartridge cases and Zeiss Axio CSM 700 confocal microscope to capture the striation marks. Principle component analysis (PCA) followed by linear discriminant analysis (LDA) were fed into a support vector machine (SVM) for classification purposes. They reported identification error rate of ~1% with 95% confidence intervals.

Traditionally in forensic science the terms class- and individual-characteristics are well used, but various interpretations and usages of these terms occur frequently in the literature. Generally, classification is considered as how results of the method are grouped. One can consider that the outcome of a method, especially in impression evidence, is a basic classification problem (match/non-match). This is in contrast to the class characteristics of the evidence.

The Association of Firearm and Toolmark Examiners (AFTE) has stated their "theory of identification" in three principles:

- The theory of identification as it pertains to the comparison of toolmarks enables opinions of common origin to be made when the unique surface contours of two toolmarks are in "sufficient agreement."
- "Sufficient agreement" is related to the significant duplication of random toolmarks as evidenced by the correspondence of a pattern or combination of patterns of surface contours. Significance is determined by the comparative examination of two or more sets of surface contour patterns comprised of individual peaks, ridges, and furrows. Specifically, the relative height or depth, width, curvature, and spatial relationship of the

---

[7] Daniel L. Cork, Vijayan N. Nair, and John E. Rolph. Some forensic aspects of ballistic imaging, Fordham Urban Law Journal, 2010.

[8] Nicholas D. K. Petraco, Helen Chan, D.Crim. Peter R. De Forest, Peter Diaczuk, Carol Gambino, James Hamby, Frani L. Kammerman, Brooke W. Kammrath, Thomas A. Kubic, Loretta Kuo, Patrick McLaughlin, Gerard Petillo, Nicholas Petraco, Elizabeth W. Phelps, Peter A. Pizzola, Dale K. Purcell, and Peter Shenkin, NCJRS Publications, 2012. https://www.ncjrs.gov/App/Publications/abstract.aspx?ID=261107 (accessed August 7th, 2014).

individual peaks, ridges, and furrows within one set of surface contours are defined and compared to the corresponding features in the second set of surface contours. Agreement is significant when it exceeds the best agreement demonstrated between toolmarks known to have been produced by different tools and is consistent with agreement demonstrated by toolmarks known to have been produced by the same tool. The statement that "sufficient agreement" exists between two toolmarks means that the agreement is of a quantity and quality that the likelihood another tool could have made the mark is so remote as to be considered a practical impossibility.

- Currently the interpretation of individualization/identification is subjective in nature, founded on scientific principles and based on the examiner's training and experience."

According to this statement the concept of sufficient agreement is achieved when the agreement firstly exceeds the best agreement demonstrated between toolmarks known to have been produced by different tools and secondly the agreement is consistent with toolmarks known to have been produced by the same tool.

Best known non-matches and ROC curves

According to an online training program funded through NIJ[9] by NFSTC and in collaboration with AFTE, the degree of correspondence that must be exceeded in order to reach sufficient agreement to effect an identification is the best known non-match (BKNM) as determined by each individual examiner and as produced by different tools. The individual examiner gains this experience during their initial training period rather than when they begin to perform their own casework examinations. Anecdotally, it is known that examiners do find better BKNMs after completion of their training[10].

In order to understand this process this concept will be elaborated upon. Given that each firearms examiner establishes their own BKNM, it seems plausible that since there are multiple examiners, there must be a range of BKNMs. This being true, the implication is that for a crime scene sample-known sample pair (CS-KN pair), we have two examiners, x and y, then each of these examiners will have their own BKNM threshold ($BKNM_x$ and $BKNM_y$ respectively). If we furthermore define the characteristics in congruence on the above-mentioned pair, the following threshold range of "sufficient agreement" can be defined as:

[9] http://www.nij.gov/training/firearms-training/module11/fir_m11_t04_05.htm accessed on 2011-11-09
[10] Brudenelle, A, personal communication

Equation 1: Variability in "sufficient agreement" between firearms examiners

$$BKNM_x < Match\ between\ CS - KN\ pair < \ BKNM_y:$$

To further understand the implications of the AFTE theory of identification, another concept needs to be addressed. This concept may be illustrated by using consecutively matching striations method (CMS) as described by Biassoti in 1959[11]. Since the introduction of this method, there has been much debate between the so-called *pattern matchers* and *line counters*.

According to Nichols, one of the frequent criticisms of the CMS method (versus the pattern matching method) is that it is too prone to false exclusions[12]. This may be restated as a non-match based on CMS when a *pattern matcher* would call it a match. Is this an unexpected behavior? In order to understand the behavior, one needs to return to the discriminating power of a method.

The discriminating ability of a method can be described by its sensitivity and specificity. The sensitivity of a method is its ability to detect a condition when the condition is present (or calling a *match*, a *match*). Specificity is the ability of a method to detect an absence when the condition is not present (or calling a *non-match* a *non-match*). The sensitivity of a method is equivalent to the true positive rate (*tpr*), whilst the specificity is equivalent to the true negative rate (*tnr*). Given the focus on methods generally used in forensic science and subject to the Daubert criteria it is important to understand the known or potential rate of error in the method. Generally courts are interested in the false positive rate (*fpr*) and their false negative rate (*fnr*) of a method.

[11] Biasotti, A.A. A statistical study of the individual characteristics of fired bullets. Journal of Forensic Sciences 4(1), 1959, 34–50.

[12] Nichols, R.G., Consecutive matching striations (CMS): its definition, study and application in the discipline of firearms and tool mark identification, AFTE Journal, 2003, 35(3), p 298- 306.

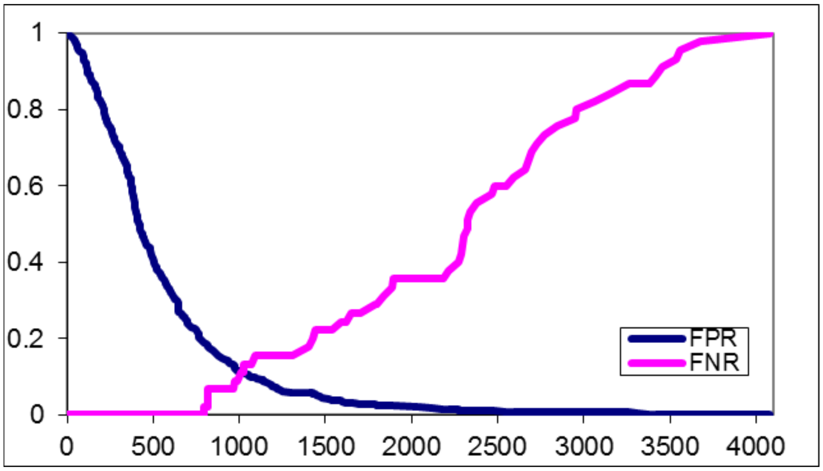

Figure 1: Generalized error curve

Let Figure 1 represent the performance of a method for the comparison cartridge cases as typically exercised in a forensic laboratory. The arbitrary measure on the x- axis represents the result of the comparison between many pairs of cartridge cases for which the ground truth is known. Figure 1 gives the *fpr* and *fnr* curves for this particular method. For each cut-off on the x-axis, a finite value for the *fpr* and *fnr* is given. The concept of a BKNM can now be defined in terms of the *fpr*. Let $BKNM_x = 1500$ and $BKNM_y = 2000$. Based on these assumptions one can consider three situations:

1. If the CS-KN pair has a score of 1000, both examiners will classify this as a non-match,
2. if the CS-KN pair has a score of 1250 then examiner *x* will classify it as a match whilst examiner *y* will classify it as a non-match, and
3. if the CS-KN pair has a score of 2500 both examiners will classify it as a match.

Let the standard (perfect) BKNM be defined as $BKNM_s$. Thus:

Equation 2: Limit of "sufficient agreement" (BKNM)

$$BKNM_i \to BKNM_s \, , fpr \to 0$$

In the example given in Figure 1, as the *fpr* tends to zero, the *fnr* tends to one. Thus, irrespective of a CMS or pattern matching approach, the concept of eliminating false exclusions is unavoidable. Conversely, if the number of false exclusions you reject increases the *fpr* rate will increase.

IBIS

The Integrated Ballistic Identification System (IBIS) developed by Ultra Electronics Forensic Technology Inc. serves as the backbone of the NIBIN system[13]. This system allows for the data-basing of images of cartridge cases and bullets. For each cartridge case there are three areas imaged, *viz.* the firing pin impression, the breech face impression, and the ejector mark. For the bullet each land-engraved area is imaged. As each new item is entered the system will search against previous entries and provides a candidate list of potential matches in the database. In order to do so the system calculates a match score of each area based on a proprietary algorithm. For cartridge cases the system will provide independent scores for each impression. Since both the firing pin and ejector in many firearms is a replaceable part, and given that their manufacture is independent of each other, each of the scores is independent. Data previously collected based on 9mm pistols are given in Figure 2[14].

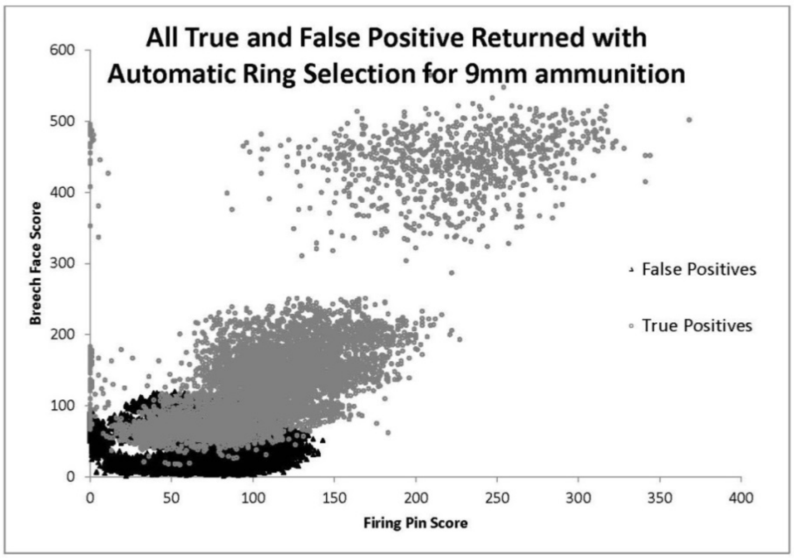

Figure 2: IBIS scores for 9mm pistols

The scores depicted in Figure 2 are from matching or same-source (grey) and non-matching or different-source (black) cartridge case pairs. There is significant overlap of some cartridge cases, yet a large number of true postives are clearly separated and this indicates the ability to call matches based on a suitable population of cartridge cases.

A simplified diagram of a small, but similar data set is provided in Figure 3.

[13] http://www.ultra-forensictechnology.com/ibis accessed 10/29/2015

[14] Scicchitano, K.M., The effect of examiner variation in cartridge case acquisition on IBIS® correlation scores and the ability of the system to return a true positive, MS thesis, West Virginia University, 2011.

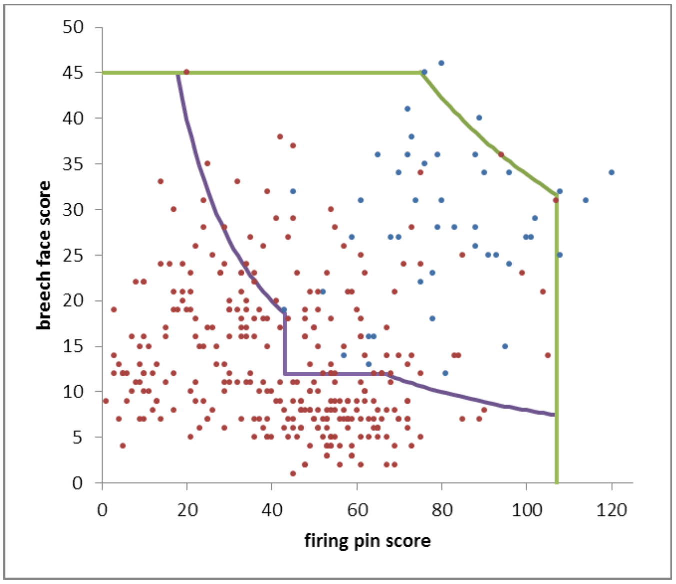

Figure 3: IBIS scores from pistols

Figure 3 can be construed as a decision space. The green curve represents the minimum boundary at which the *fpr* = 0. Any comparison that yields a value beyond this boundary is a true positive. The purple curve represents the maximum boundary at which the *fnr* = 0. Each curve is comprised of three lines. The vertical straight sections are for the firing pin score, the horizontal straight lines are for the breech face score, and the hyperbola represents the product of the firing pin and breech face scores.

The investigation involved the acquisition of cartridge cases fired by a number center-fire handgun calibers (including two carbines and one rifle in popular handgun calibers) typically found in crime scene work. Each firearm was used to shoot 100 rounds. The fired cartridge cases were entered in to the Heritage IBIS system (2D system) in order to generate the match data. The acquisition method to be followed will follow guidelines established in a previous study[15]. The data will be mined to evaluate the various within/between relationships such as calibers, model, makes, firing conditions, etc. An overall evaluation of the efficacy of the method and the necessary Daubert requirements will be provided[16]. The data generated were transformed to develop likelihood ratios for the interpretation of firearms evidence.

Bayesian networks

Bayesian networks (BNs) were used in the project to provide a framework for interpretation of the collected data. Equation 3 is the odds form of Bayes' theorem.

[15] Scicchitano, K.M., The effect of examiner variation in cartridge case acquisition on IBIS® correlation scores and the ability of the system to return a true positive, MS thesis, West Virginia University, 2011.

[16] Daubert v. Merrell Dow Pharmaceuticals (92-102), 509 U.S. 579 (1993).

Equation 3: Odds form of Bayes' theorem

$$\frac{Pr(H_1|R,I)}{Pr(H_2|R,I)} = \frac{Pr(R|H_1,I)}{Pr(R|H_2,I)} \times \frac{Pr(H_1|I)}{Pr(H_2|I)}$$

In Equation 3, H_1 and H_2 are the competing hypotheses, *R* is the evidence, and *I* is the background information. The left hand side of this equation is the posterior odds. The first term on the right hand side is the likelihood ratio (LR) (also symbolized by *V*), and the second term is the prior odds[17]. As an example, the numerator in the LR can be read as "the probability of the evidence, R, given hypothesis H_1 and the background information, I." In the normal evaluation of forensic evidence the likelihood ratio is defined as the ratio of "the probability of the evidence given that the accused committed the crime divided by the probability of the (same) evidence given that the crime was committed by some other person other than the accused."

The hypothesis node at the root of the network (Match) is based on the prosecutorial hypothesis (the accused committed the crime, H_p, same gun hypothesis, or Match =Yes) and the defense hypothesis (someone else, other than the accused committed the crime, H_d, different gun hypothesis, or Match = No). The ratio $\frac{H_p}{H_d}$, before any evidence is applied to the Bayesian network, will result in the prior odds. After the evidence is applied to the Bayesian network, the ratio $\frac{H_p}{H_d}$ will result in the posterior odds. The quotient of the posterior odds and the prior odds will give the likelihood ratio.

Basic network structures

Bayesian networks are a type of graph known as directed acyclic graphs (DAGs). It is specifically comprised of nodes (variables) each having two or more states (a state is a condition that a variable could assume). Nodes are connected by edges (arcs or arrows). The edges indicate the direction of the conditional probabilities.
All Bayesian networks are combinations of three basic node structures, *viz.* serial, converging, and diverging. Each structure will be defined in turn.

Serial nodes: $\boldsymbol{A} \rightarrow \boldsymbol{B} \rightarrow \boldsymbol{C}$

Initially no nodes are instantiated (i.e. no state assigned to the variable). If evidence is found for node A or node C, then instantiation of that node will affect the other two nodes. However, if evidence is known for node B is found, then that state of node B can be instantiated. Once node B is instantiated, then the state of node A will have no effect on node C, since node B blocks communication from node C to node A (and vice-versa).

[17] Taroni, F, Aitken, C, Garbolino, P, Biedermann, A, Bayesian Networks and Probabilistic Inference in Forensic Science (Statistics in Practice), Wiley, 2006.

Converging nodes: $\boldsymbol{A} \rightarrow \boldsymbol{C} \leftarrow \boldsymbol{B}$

If node C is not instantiated (no evidence available) then node A and node B are independent of each other. In other words, instantiating node A will have no effect on node B. When node C is instantiated then it “opens up” flow between node A and node B. If evidence from node A can be used to explain node C, then node B has less influence on node C, since node A and node B are “competing” explanations for node C.

Diverging nodes: $\boldsymbol{A} \leftarrow \boldsymbol{B} \rightarrow \boldsymbol{C}$

Assuming that we are uncertain about node B, if we get evidence for node A, then this evidence changes the probabilities of the states in node B, which in-turn changes the probabilities of the states in node C. However, if evidence for node B is found, a particular state can be instantiated. Subsequently, any change in node A will have no effect on node C.

Table 1: Nodes developed for inclusion in the Bayesian networks with a description and an example

Field	Description	Example
CaseID_Sample	String as a unique descriptor of each sample test fire. Includes info regarding the firearm, ammunition, primer, exhibit number	*CBN-BB-UNK-1001-0001*
ExhibitNumber_Sample	Exhibit number of the sample cartridge case	*1*
Rank	Rank position of the firing pin score	*3*
CaseID_DB	As for *CaseID_Sample* but for the cartridge case in the database returned list comparison	CBN-BB-UNK-1001-0015
ExhibitNumber_DB	Exhibit number of the database cartridge case	*15*
BF	Breech face score of the sample and database cartridge cases	*146*
FP	Firing pin score of the sample and database cartridge cases	*157*
Match	Status of the sample and database cartridge cases whether they originate from the same firearm	*Yes*
Make_DB	Make of the firearm which discharged the cartridge case from the database	*SCCY*
Model_DB	Model of the firearm which discharged the cartridge case from the database	moCPX
Ammo_DB	The make of ammunition fired by the database firearm	Blazer
Caliber_DB	The caliber of the database firearm	9mm
Firing_Pin_Type_DB	The type of firing pin formed on the database cartridge case by the database firearm	Circular
Make_Sample	Make of the firearm which discharged the sample cartridge case	*SCCY*
Model_Sample	Model of the firearm which discharged the sample cartridge case	moCPX
Ammo_Sample	The make of ammunition fired by the sample firearm	Blazer
IdentifierGun_Sample	String as a unique descriptor of each sample firearm. Contains the last 5 characters of the serial number prefaced by an "X". In the case of leading zeros, these are also replaced by an "X"	X97569
IdentifierGun_DB	String as a unique descriptor of each database firearm. Contains the last 5 characters of the serial number prefaced by an "X". In the case of leading zeros, these are also replaced by an "X"	X97569
Caliber_Sample	The caliber of the sample firearm	9mm
Firing_Pin_Type_Sample	The type of firing pin formed on the database cartridge case by the sample firearm	Circular
Type_Sample	The type of sample firearm such as revolver, pistol, or carbine	Pistol
Primer_Sample	The manufacturer of the primer used in the reloading of the sample cartridge case	Unknown
Primer_DB	The manufacturer of the primer used in the reloading of the database cartridge case	Unknown
Drag_Mark_Sample	The presence of a drag mark on the sample cartridge case	Yes
Drag_Mark_DB	The presence of a drag mark on the database cartridge case	Yes
Reload	Whether or not the sample cartridge was reloaded	No
Rank_BF	Rank position of the breech face score	*1*
BFFP	The product of the breech face and firing pin scores	22922
CaseID_pre	The 1st 3 characters of the *CaseID_Sample* string	CBN
BF_norm	Normalized breech face score	10.605288
FP_norm	Normalized firing pin score	6.797851
BFFP_norm	Normalized product of breech face and firing pin scores	22.117015
Same_Model	Whether or not the *Model_Sample* and *Model_DB* are the same	Yes
ActionLB_Sample	Whether or not the *Model_Sample* is of a locking breech action	Yes
ActionLB_DB	Whether or not the *Model_DB* is of a locking breech action	Yes

Data Acquisition and Processing

The general process for acquisition of a sample on the IBIS system is as follows: A *case file* is created which can contain several exhibits (samples). A range of information, such as the investigator, case number, offence type, etc., is contained within this *case file*. The data contained within the *case file* is not easily accessible by the examiner from IBIS. In order to relate the data to a particular set of scores, the data were encoded into the *case file* identifier. The following string is an example of a *case file* identifier: "AAN-UK-SSG-0313-0901." The "AAN" refers to a particular firearm; the first letter defines the make of the firearm (in the case: "A"=Arcus), the second letter is a letter specific to the test set, and the third, "N", indicates the caliber (in this case: "N" = 9mm Luger). The "UK" refers to the make of ammunition; in this case it is “unknown” since this cartridge was reloaded. The “SSG” relates to the reloading data. The first “S” indicates that a small pistol primer was used; the second “S” indicates that the manufacturer of the primer was Sellier and Bellot, and the “G” indicates that Hodgdon TiteGroup smokeless powder was used. The “0313” indicates that the test fire took place during March of 2013 and the “0901” is a unique identifier for the particular cartridge case (ExhibitNumber).

Since IBIS will only compare between and not within *case files*, a new case file is created for each cartridge case. Upon submission to the IBIS database a correlation will be performed. The report (see **Error! Reference source not found.** for an example of the IBIS correlation report) needs to be processed in order for the data to become amenable to analysis. In the header, the Case ID will become the CaseID_Sample in the data file. Note in the report “Test ordered by Firing Pin” is the Rank value for firing pin scores. The data from the report that are required, in addition to the CaseID_Sample, the Rank, Case ID, Exhibit Number (although redundant in the Case ID column), Breech Face, and Firing Pin columns. The Case ID in the column will become the CaseID_DB in the final data file. The CaseID_Sample will be posted to each entry since the scores are the result of the comparison of the sample cartridge case (CaseID_Sample) against the particular database sample (CaseID_DB). R via RStudio is used to clean up, rearrange, and expand the data into the final format. For both the sample and database cartridge cases, items such as same gun/different gun status (Match), firearm make, model, serial number, firing pin type[18], breech action[19], firearm type[20], and caliber[21], as well as ammunition manufacturer, primer manufacturer, presence of a drag mark[22], reloaded ammunition status[23], and whether the sample and database cartridge cases were fired by a firearm of the same make and model, are all introduced into the dataset. The breech face rank is also determined.

The data from the IBIS correlation reports (see Figure 4) are processed using the script (see Appendix A on page 233) to form a *.CSV file. The data in a typical *.CSV file are given in Table 1. The variables in Table 1 are then used as the nodes developed for inclusion in the Bayesian networks.

The *.TXT report file is processed as follows. The report in Figure 4 will be used as an example for explanation. This report is for the correlation of the cartridge case RNN-UK-SFG-0320-0501 against the database. The caliber of this cartridge case is 9 mm Luger (Caliber: 9LG*). The

[18] Circular, Glock-type, etc.
[19] Blowback, recoil, etc.
[20] Rifle, carbine, pistol, revolver.
[21] Important, for example, in .38 Special and .357 Magnum cartridge case comparisons.
[22] The presence or absence of a drag mark has to be added after visual inspection of all of the images in IBIS. There is no way of extracting this data since it is not determined.
[23] Yes or no – reloaded or factory bought ammunition.

number of returned comparisons is 2009 (Sample size). The listed data are the results for each comparison. Using the second entry as an example, it can be seen that: Rank: 2, Case ID: RFN-UK-SSG-0320-0391, Exhibit Number: 0391, Site Name: MDEMO4-DAR, Breech Face: 67, Firing Pin: 226, and EM or RF: 0. As discussed above, each of these entries will be processed through the script to create a data file. For the RNN-UK-SFG-0320 set, there will be 100 files similar to this one (RNN-UK-SFG-0320-0501 through RNN-UK-SFG-0320-0600). All of the RNN files provide 9,745[24] match (same gun entries) and 78,654 non-match (different gun entries) with an average of 884 entries per cartridge case submitted to the database. For this particular firearm (a Ruger P95), a number of different test fires were collected (RVN, RFN, and RPN[25]) in addition to RNN.

The Bayesian networks were developed by using the integrated learning algorithms[26] together with logical constraints from the datasets. The networks can be built in graphical user interface of the software (Netica™) or through an R package called RNetica.[27] In all of these instances, the root node is the Match node. This is what the analyst would want to assess in order to provide an interpretation of the evidence. The script used to create the Bayesian networks is given in Appendix B (page 250).

---

[24] This is slightly less (155) than the maximum possible returned scores of $2 \times {}^{100}_{2}C = 9900$. The number is doubled because all submissions to IBIS were re-correlated.

[25] The difference between these submissions was primer manufacturer -- RVN (TulAmmo), RFN (Sellier & Bellot), RPN (Remington), and RNN (Federal).

[26] See for example Friedman N., Geiger D., and Goldszmidt M., Bayesian network classifiers, Machine Learning, 29, 1997, 131-163.

[27] Almond R., R interface to Netica® Bayesian Network Engine, Version 0.4-4, 2015/06/29.

IBIS Correlation Results

Reference Case Information

Case ID: RNN-UK-SFG-0320-0501
Site Name: MDEMO4-DAR
Law Agency: (Unknown LAW Agency)
Event: (Unknown)
Comment:

Reference Exhibit Information

Exhibit Number: 0501
Event: (Unknown)
Caliber: 9LG*
Acq. Person: EXAMINER
Comment:

Sample Size 2009
Tests ordered by ***Firing Pin***

Rank	Case ID	Exhibit Number	Site Name	Breech Face	Firing Pin	EM or RF
1	RNN-UK-SFG-0320-0509	0509	MDEMO4-DAR	117	226	0
2	RFN-UK-SSG-0320-0391	0391	MDEMO4-DAR	67	226	0
3	RNN-UK-SFG-0320-0503	0503	MDEMO4-DAR	80	213	0
4	RFN-UK-SSG-0320-0311	0311	MDEMO4-DAR	57	212	0
5	RFN-UK-SSG-0320-0339	0339	MDEMO4-DAR	83	210	0
6	RFN-UK-SSG-0320-0385	0385	MDEMO4-DAR	47	203	0
7	RFN-UK-SSG-0320-0370	0370	MDEMO4-DAR	52	201	0
8	RFN-UK-SSG-0320-0333	0333	MDEMO4-DAR	67	199	0
9	RFN-UK-SSG-0320-0358	0358	MDEMO4-DAR	51	197	0
10	RFN-UK-SSG-0320-0379	0379	MDEMO4-DAR	58	195	0
11	RFN-UK-SSG-0320-0312	0312	MDEMO4-DAR	53	194	0
12	RNN-UK-SFG-0320-0505	0505	MDEMO4-DAR	56	193	0
13	RFN-UK-SSG-0320-0329	0329	MDEMO4-DAR	52	190	0
14	RNN-UK-SFG-0320-0504	0504	MDEMO4-DAR	80	189	0
15	RFN-UK-SSG-0320-0357	0357	MDEMO4-DAR	12	186	0
16	RFN-UK-SSG-0320-0334	0334	MDEMO4-DAR	43	186	0
17	RFN-UK-SSG-0320-0364	0364	MDEMO4-DAR	33	184	0
18	RFN-UK-SSG-0320-0374	0374	MDEMO4-DAR	17	183	0
19	RFN-UK-SSG-0320-0349	0349	MDEMO4-DAR	101	183	0
20	RFN-UK-SSG-0320-0307	0307	MDEMO4-DAR	46	182	0
21	RFN-UK-SSG-0320-0306	0306	MDEMO4-DAR	61	182	0
22	RFN-UK-SSG-0320-0384	0384	MDEMO4-DAR	13	181	0
23	RFN-UK-SSG-0320-0345	0345	MDEMO4-DAR	20	181	0
24	RFN-UK-SSG-0320-0323	0323	MDEMO4-DAR	83	181	0
25	RFN-UK-SSG-0320-0372	0372	MDEMO4-DAR	36	179	0
26	RFN-UK-SSG-0320-0355	0355	MDEMO4-DAR	80	179	0
27	RFN-UK-SSG-0320-0363	0363	MDEMO4-DAR	44	178	0
28	RFN-UK-SSG-0320-0351	0351	MDEMO4-DAR	51	178	0
29	RNN-UK-SFG-0320-0508	0508	MDEMO4-DAR	92	177	0
30	RFN-UK-SSG-0320-0395	0395	MDEMO4-DAR	59	177	0
31	RFN-UK-SSG-0320-0387	0387	MDEMO4-DAR	20	177	0
32	RFN-UK-SSG-0320-0308	0308	MDEMO4-DAR	68	175	0
33	RFN-UK-SSG-0320-0304	0304	MDEMO4-DAR	49	172	0
34	RFN-UK-SSG-0320-0356	0356	MDEMO4-DAR	60	171	0
35	RFN-UK-SSG-0320-0371	0371	MDEMO4-DAR	66	170	0
36	RFN-UK-SSG-0320-0360	0360	MDEMO4-DAR	61	170	0
37	RFN-UK-SSG-0320-0362	0362	MDEMO4-DAR	60	169	0
38	RFN-UK-SSG-0320-0338	0338	MDEMO4-DAR	56	169	0
39	RNN-UK-SFG-0320-0510	0510	MDEMO4-DAR	119	168	0
40	RFN-UK-SSG-0320-0394	0394	MDEMO4-DAR	58	168	0
41	RFN-UK-SSG-0320-0389	0389	MDEMO4-DAR	22	168	0
42	RFN-UK-SSG-0320-0328	0328	MDEMO4-DAR	67	168	0
43	RFN-UK-SSG-0320-0380	0380	MDEMO4-DAR	71	166	0
44	RFN-UK-SSG-0320-0390	0390	MDEMO4-DAR	56	165	0
45	RFN-UK-SSG-0320-0336	0336	MDEMO4-DAR	56	165	0
46	RFN-UK-SSG-0320-0310	0310	MDEMO4-DAR	42	165	0
47	RFN-UK-SSG-0320-0366	0366	MDEMO4-DAR	65	162	0
48	RFN-UK-SSG-0320-0318	0318	MDEMO4-DAR	69	162	0
49	RFN-UK-SSG-0320-0314	0314	MDEMO4-DAR	70	159	0
50	RFN-UK-SSG-0320-0400	0400	MDEMO4-DAR	49	158	0

For official use only
Sensitive but unclassified

05/21/2013 *Pages 1/7*

Figure 4: First page of a correlation report generated by IBIS

Analytical Approach

This project was an attempt to provide a broad base analysis of firearms evidence through a statistical approach mainly utilizing Bayesian networks. Part of the approach included a mathematical articulation of the AFTE theory of identification. In any interpretation of analytical data in general and forensic analyses in particular, needs a clear understanding of the variability of any measurement. The random nature of production processes and the associated wear-and-tear of mechanical devices play an important role in the analysis of firearms evidence. The main challenge to the interpretation of firearms evidence is the general lack of numerical data. This challenge remains one of the greatest facing forensic science today. However, there does need to be a clear delineation between the roles of comparison and evidence evaluation. It is clear that most, if not all, of forensic sciences require a comparative component. Many will point to DNA analysis as the gold standard for forensic scientists. A cursory assessment of DNA analysis requires that the analyst perform a comparison of a DNA profile recovered from an item of evidence at a crime scene with that from a known source. A drug chemist will compare a mass spectrum of a white powder recovered from suspect with that of a mass spectrum from a cocaine standard. The questions that need to be answered are how much of a difference between the known and the questioned sample can be tolerated in order to assign a measure of similarity between the two.[28] This provides us with a basis for considering the elements of this project. A major challenge is the determination of a suitable sample size to be able to describe the variability in a particular sample type. For firearms, this question may be posed as "how many cartridge cases are required to describe the variability present in markings on the cartridge case?" Other examples of this problem are how many fingerprints of a particular finger are needed to describe the variability in distortion of that particular fingerprint. This project does not intend to give a complete answer to this problem since the extent of the problem is far beyond its scope.

The IBIS system was evaluated in respect of its performance in classifying cartridge cases fired by a variety of firearms in 9 mm Luger. The breech face and firing pin scores, as well as their product, were used as classifiers in this respect. A typical binary classification system was used to determine false positive- and false negative-rates as well as the area under the receiver operating characteristic curve[29]. These types of rates are those which are classically called for in the Daubert criteria.

In order to assess the reliability of the IBIS system used in this study, a repeatability and reproducibility study was performed. The focus of this part of the study was to assess the effects of user interaction and system performance. Additionally, an evaluation of the NIST standard cartridge case was performed to assess the limits of performance of the IBIS system and the expectation in variability of comparisons.

It was initially hypothesized that the IBIS results could be used to classify the firearm make a model from a questioned cartridge case. All attempts in this regard proved unsuccessful.

---

[28] This measure may be considered to be the "same", "a match" , or other similar name.

[29] See, for example, Bradford T. Ulery B.T., Hicklin R.A. , Buscaglia J., and Roberts M.A., Accuracy and reliability of forensic latent fingerprint decisions, PNAS, 108(19), 2011, 7733–7738.

In an attempt to improve the discriminating power of the IBIS system it was suggested, by the manufacturer, that the data generated be normalized[30]. Improvements in discrimination for classification system will have a positive effect. There was a small improvement in the results, but the normalization parameters were not consistent across the various firearm makes and models. The approach has some benefit for general classification, but requires that the known non-match distribution be specified, *ab initio*. This is obviously not possible for the samples typically encountered in the forensic laboratory.

To assess other possible classification schemes, an analysis of the 9 mm data was performed using generally applied machine learning techniques. When using the breech face scores, the firing pin scores, and their product the match accuracy for ranges between 52% (generalized linear model) and 62% (k-nearest neighbors and decision trees) while the worst non-match accuracy was 94.60% (k-nearest neighbors).

The manufacturers of the IBIS system introduced a 3D system which has been widely implemented. To ensure compatibility of results, a test set was submitted to both system types. The firearms were selected based on their performance on the 2D IBIS system used in the study. The major advantage of the newer IBIS system is the ability to correlate the side-light images of the cartridge cases. A view to which firearms examiners are more accustomed.

The program managers from the Defense Forensic Science Center (DFSC), Office of the Chief Scientist provided a number of sample sets (questioned and known cartridge cases) to assess the performance of the Bayesian networks created during the study. In all cases the sets were submitted in a blind fashion. After analysis the ground truth data were provided to analyze the results. There were two main sets (USACIL and Baldwin). The first set was one made by the DFSC and represented a variety of firearms. The second set was of cartridge cases fired from 25 Ruger SR9 pistols. Cartridge cases from these pistols were already in the database. Results were analyzed by initially excluding all of the related data, and then by including it. Samples from this second set had previously been examined by volunteer firearms examiners. The results of their examinations were only released to this study after the tests results were finalized.

What is a "match"?

To ensure clarity in further discussion of this project it may be useful to provide some working definitions of common terms. The term "match" in forensic science, generally means that some item of evidence is attributed to a particular source. This process generally implies that the true state is unknown. In firearms examination, in particular, match generally implies a same gun source attribution whereas a non-match implies an attribution of different sources for the cartridge case and firearm (different gun). Matches between objects of the same class are usually

[30] A conventional approach used to transform data before analysis. Oftentimes used as a preprocessing technique in principal component analysis (PCA).

achievable when the within class variability (intra-variability) is significantly smaller that the between class variability (inter-variability).

Equivalent terms for match may be identification and individualization. Common source and identity must be understood within the framework of samples. It is known that two things cannot be numerically identical, or simply stated two things cannot be one thing. If two "objects" have a sufficient quantity of characteristics in common and such characteristics are relatively rare in the general population, then one can think of a match being obtained. A classic example is that of DNA where, if the alleles of each locus are the same between a known and questioned sample then a "match" is achieved. The value of this match is based on the rarity of those allele combinations at each locus. A random match probability can be generated to estimate this rarity.

When evaluating the ability of a technique to discriminate a "gold standard[31]" is typically used. This standard is the technique or method used to determine the outcome of a test. Thus if we have a new test it can be evaluated against the gold standard to determine its efficacy.

Source attribution of produced surrogate evidence is perhaps a good descriptor of cartridge case comparisons. The single object is the firearm from which the two cartridge cases originate. The examiner will use the cartridge case from the crime scene as the unknown and the test fired cartridge as the known. Upon evaluation of the feature a match status is inferred. Two cartridges fired from a single gun are said to match, by means of deductive reasoning, when the induced "law" of firearms identification is invoked.

If m cartridges are fired from firearm M and n cartridges from firearm N, then cartridge m_i matches m_j (as does n_i and n_j because of their common source) but cartridge m_i does not match cartridge n_i. This is true irrespective of how many features an examiner may or may not find. The match status in examinations is based on features.

A further feature that is important for sustained comparison can be borrowed from fingerprint examination and is permanence/persistence. The term persistence is preferred since, in fingerprints, it implies that features do not change unless some major deformation takes place. Any major deformation result in the regeneration of the deformed feature itself through the normal biological process and the new feature becomes persistent. In firearms such major deformations may take place (a part is replaced or a part receives major damage). There is also normal wear and tear in a firearm which may present itself as gradual changes to the surfaces of the firearm that contact the various components of a cartridge.

[31] A gold standard is that which is generally accepted to give the "true" answer.

Inter- and Intra-variability

Variability is probably the critical area in forensic science that requires attention as it pertains to evidence interpretation. This is especially difficult since, in many cases, forensic scientists and firearms examiners in particular have a sample size of one. In firearms examination, observation of bullets and cartridge cases fired by a single firearm will result in characteristic markings which are similar in structure between the cartridge cases, as an example, of successive shots, and may be consistent across series of many shots, or maybe inconsistent between successive shots through a series of shots. Being able to discern the similarities and differences is integral to the understanding of firearms examination and in the training of firearms examiners.

The chief concern is how one may quantify this variability. The findings examiner is faced with two distinct problems in terms of variability. The first relates to the match (same gun) proposition, and the second to the non-match (different gun) proposition. In the section entitled Bayesian networks (page 13), the idea of consideration of two different propositions which are mutually exclusive and exhaustive was introduced. Let us consider the situation that a typical firearms examiner faces during the comparison of a questioned and known cartridge case. She has to determine whether or not a particular firearm (the source of the known cartridge cases) discharged the questioned cartridge case. Generally the propositions that are faced in court are: (a) the firearm in question is the one that discharged the cartridge case, and (b) the firearm in question is not the one that discharged the cartridge case but some other firearm. Let us assume that the firearms examiner has these cartridge cases under inspection in the comparison microscope. That which is being observed by the examiner is the evidence[32] and the firearms examiner needs to assess two aspects: (1) If I accept that the suspect firearm fired both the questioned and the known cartridge case, what is the probability that I would observe this set of features present in the comparison of the two? (2) If I accept that the suspect firearm did not fire the questioned cartridge case but did fire the known cartridge case[33], what is the probability that I would observe this set of features present in the comparison of the two? Each of these probabilities is based the knowledge and experience of the examiner in comparing known pairs of cartridge cases from the same gun and from different guns. Such examinations will result in an assessment of the variability in such comparisons. If the variability is high in cartridge cases from the same firearm, then it will be difficult to differentiate that firearm from other firearms.

The second problem of variability is described in the following data. This problem may be explained by way of an example.[34] Imagine a set of cartridge cases which will discharged by the same firearm. Many factors may influence the quality of the breach face and firing pin impressions. As a result this set of cartridge cases may be considered to be representative of the cartridge cases fired by this particular firearm. Assume now, that this firearm was used as a

[32] In this context, the evidence is the act and outcome of the comparison rather than the physical evidence itself (i.e. the known and questioned cartridge cases).
[33] The firearms examiner most likely fired the known cartridge cases themselves.
[34] All others factors being equal.

weapon during a crime. During the crime a single shot was fired and a cartridge case was left at the crime scene. After some investigation a suspect is developed and a firearm is seized. The firearms examiner receives both the suspect firearm and the questioned cartridge case. The firearms examiner now fires say three test fires (known cartridge cases). The questioned cartridge case can be considered to be a single sampling from the distribution of all cartridge cases fired by the firearm. The known cartridge cases are also a sampling of the full distribution. If the knowns represent the "average" cartridge case and the questioned cartridge case comes from one of the tails of the distribution, then it is possible that a firearms examiner may not be able to affect an identification because of the nature of the questioned sample (a false negative result). It is critical, therefore, that the firearms examiner has a good understanding of the variability of cartridge cases fired by a particular firearm. This will also allow the examiner to determine a suitable sample size for the test fires since firearms examiners do not have control over questioned cartridge cases but they do have control over the generation of known cartridge cases. This situation is further supported by the results of the analysis of the NIST standard cartridge case (see page 70). The NIST standard cartridge cases are specifically produced to eliminate as much variability as possible. Figure 79 illustrates that the separation of the matches from the non-matches is easily achieved when no variability is present in the structure of the cartridge case. This is certainly not the case in successive shots fired by the same firearm.

Comparison and identification is thus ultimately dependent upon the intra-variability and inter-variability of cartridge cases from various firearms. Two new[35] .45 ACP caliber pistols (Glock 21 Gen4 and Taurus 24/7 G2) were used to fire 50 cartridges of the same brand (Federal American Eagle).These two pistols were chosen to represent their class types.   Each cartridge was marked with a permanent marker to identify it as the n^{th} shot fired through the firearm. This seemingly simple inter-cartridge compassion becomes quite complex. Figure 5, Figure 6, Figure 7, and Figure 8 illustrate the breech face and firing pin scores comparison of 50 successive shots of $shot_{n+1}$ versus $shot_n$.

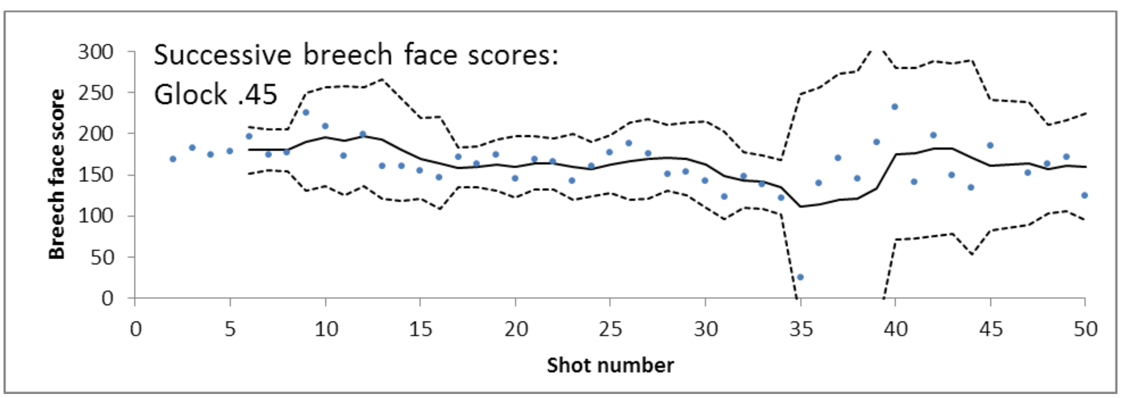

Figure 5: Successive breech face scores of 50 cartridge cases fired in a Glock 21 (.45 ACP)

[35] In this study a new firearm is that which has been purchased as new from a dealer (it may have been fired as part of the production process).

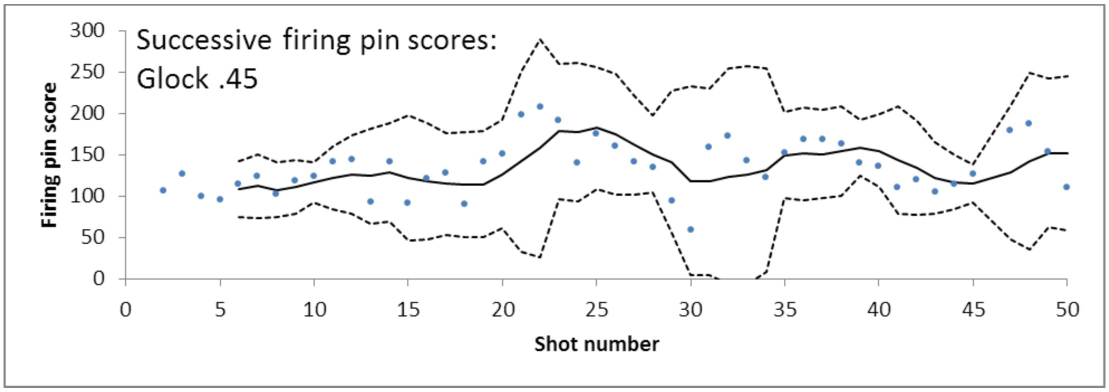

Figure 6: Successive firing pin scores of 50 cartridge cases fired in a Glock 21 (.45 ACP)

In Figure 5, Figure 6, Figure 7, and Figure 8 the breech face and firing pin scores were obtained from the IBIS system starting with $shot_2$ versus $shot_1$, then $shot_3$ versus $shot_2$, etc. A rolling average of five scores was used to evaluate the trend of changes in score. The confidence intervals were determined using a 95% confidence level and a *t*-distribution with four degrees of freedom. It appears as if there is no change in the scores even though there is a large variation in the confidence intervals between shots.

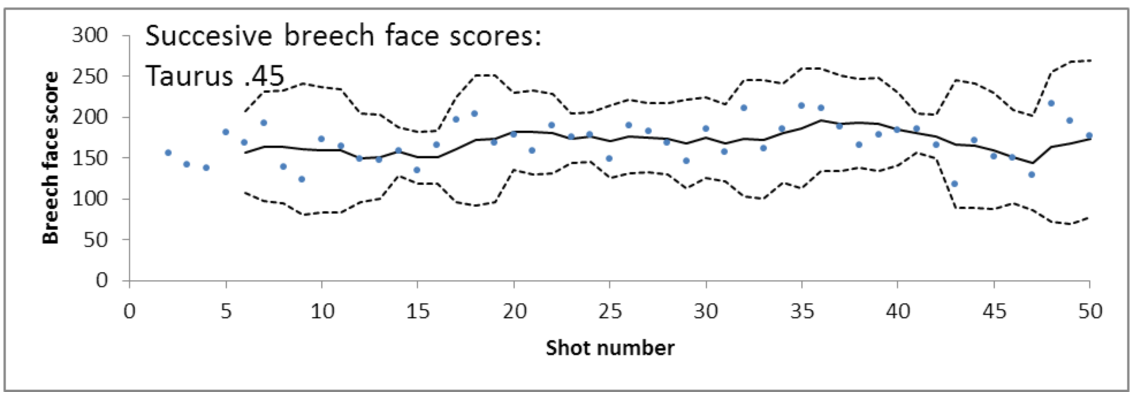

Figure 7: Successive breech face scores of 50 cartridge cases fired in a Taurus 24/7 G2 (.45 ACP)

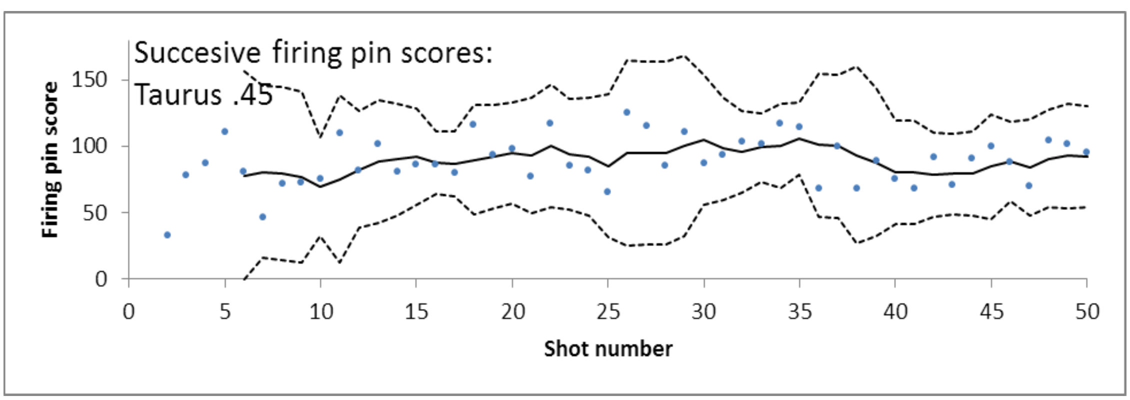

Figure 8: Successive firing pin scores of 50 cartridge cases fired in a Taurus 24/7 G2 (.45 ACP)

When multiple shots from the same firearms are entered in the IBIS system, one would expect that the cartridges previously entered from the same firearm would feature high in the generated candidate list. Thus with each additional entry, *p*, then *p-1* candidates would be expected in the candidate list for the particular firearm.

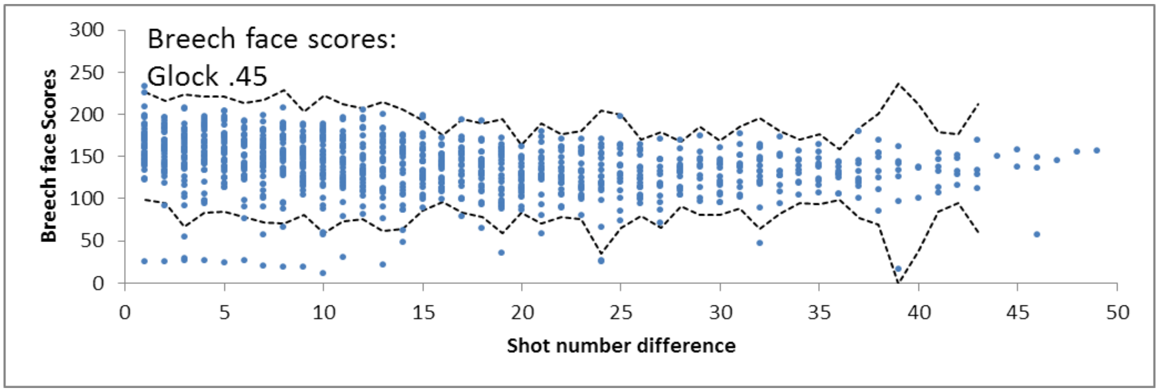

Figure 9: Variation of breech face scores by separation for the 50 Glock 21 Gen 4 cartridge cases

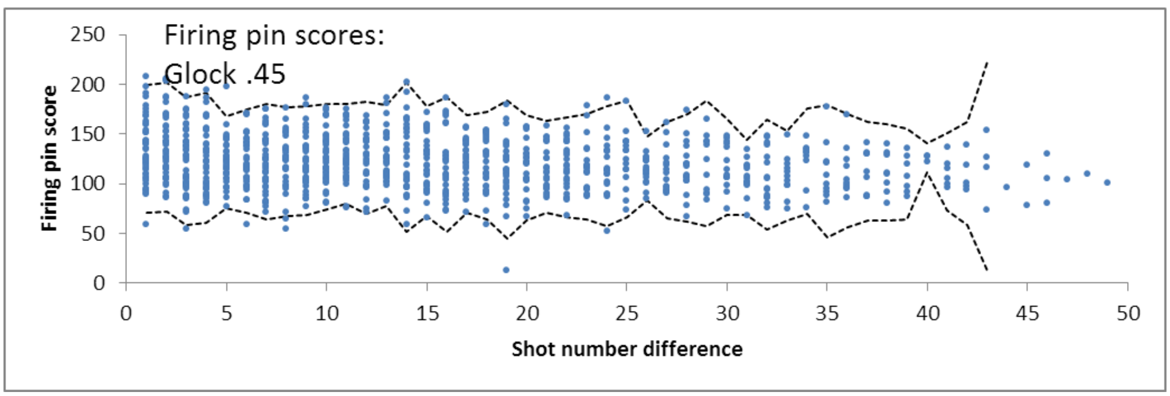

Figure 10: Variation of firing pin scores by separation for the 50 Glock 21 Gen 4 cartridge cases

In Figure 9 and Figure 10, the abscissa indicates the separation of a comparison (shot number difference). This can be illustrated as follows: Shot $_{22}$ and shot$_{21}$, and shot$_{44}$ and shot$_{43}$ are separated by one shot. Shot$_{33}$ and shot$_{25}$, and shot$_{48}$ and shot$_{40}$ are separated by eight shots. At each separation all of those scores are indicated. Ideally[36] at a separation of one, one would expect 49 values, at separation two, one would expect 48 values, and so on. The dotted lines represent the 95% confidence interval for each separation accounting for sample size. No intervals are plotted for the last few since, for example, for a separation of 49 there can only be one instance (shot$_{50}$ versus shot$_{1}$). Evaluating these results for the Glock .45 ACP caliber pistol would seem to suggest that there is no change between the various separations although the distributions within each separation are relatively large.

[36] All comparisons were returned in the IBIS candidate list.

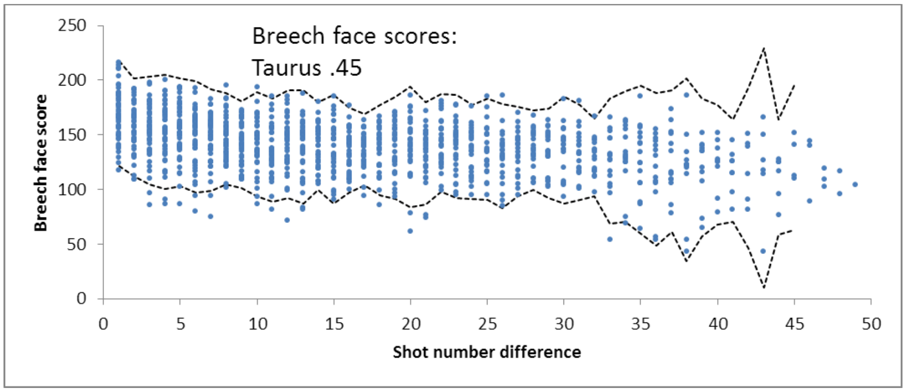

Figure 11: Variation of breech face scores by separation for the 50 Taurus 24/7 G2 cartridge cases

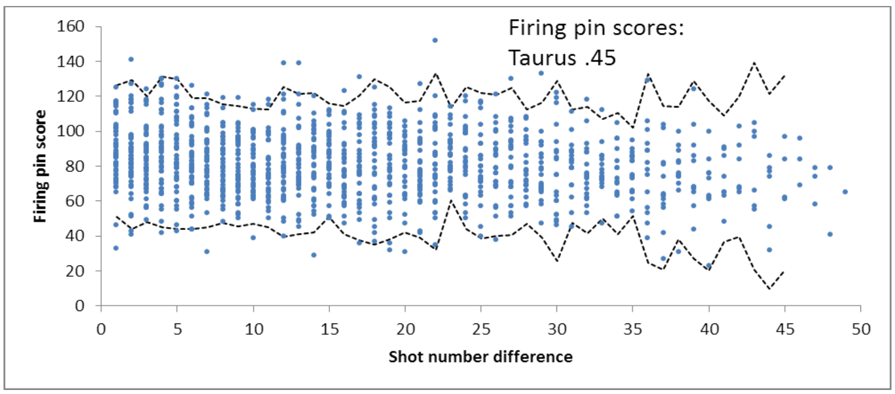

Figure 12: Variation of firing pin scores by separation for the 50 Taurus 24/7 G2 cartridge cases

Figure 11 and Figure 12 provide the shot separation plots for the .45 ACP Taurus 24/7 G2 pistol which was used in this test. The performance of the Breech face scores is similar to that of the Glock, but the Firing In scores are more spread out and at a lower mean score that the Glock pistol.

Figure 13 provides a plot for the product of the breech face and firing pin scores for the Taurus 24/7 G2 in .45 ACP (it is assumed that the scores of the breech face and the firing pin are independent and that these scores can be combined in a multiplicative fashion).

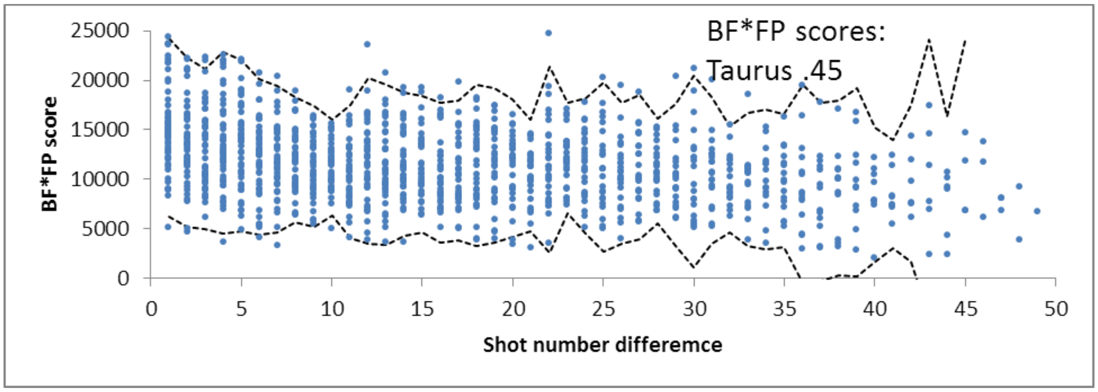

Figure 13: Variation of the product of breech face and firing pin scores by separation for the 50 Taurus 24/7 G2 cartridge cases

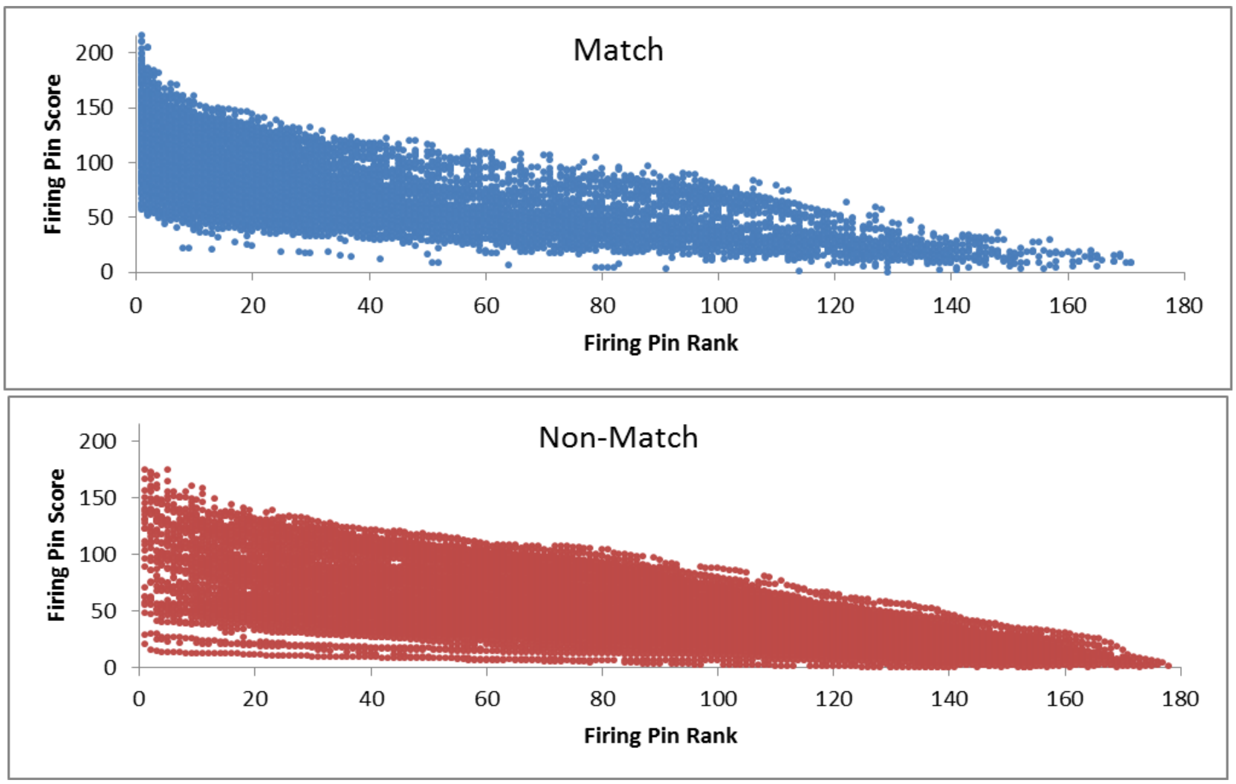

Figure 14: Plot of firing pin scores by firing pin rank and Match status

In Figure 14, the match and non-match firing pin scores are plotted against the firing pin rank for all firearms and ammunition types. As would be expected, there is a general decrease in score with an increase in rank. At high rank (low numerical values), the match scores have a higher minimum than the non-matching scores. The match scores have a higher density at higher ranks. The reverse is true for the non-match scores.

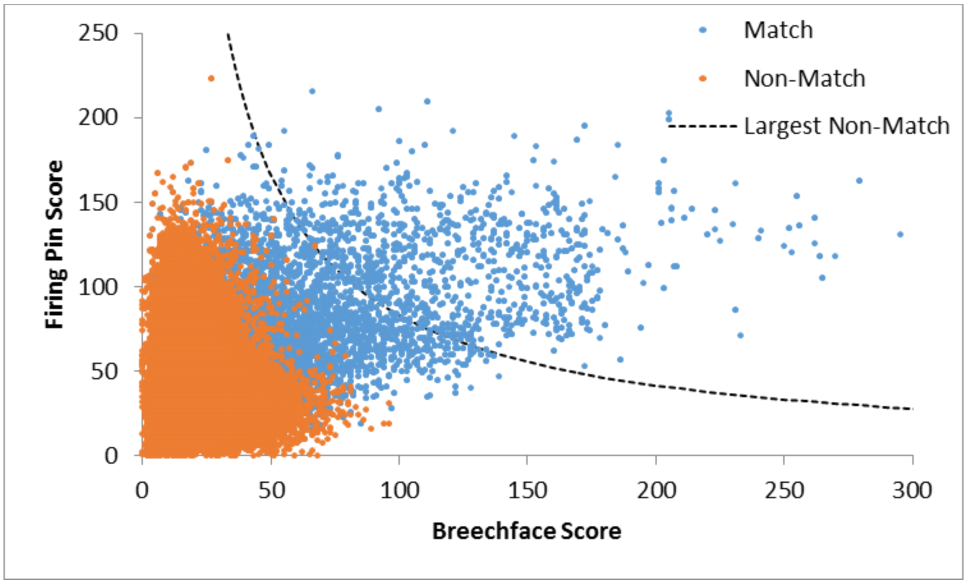

Figure 15: Firing pin versus breech face scores, by match. The best known non-match lines (Largest non-match) is indicated

All of the data from these comparisons were included in Figure 15. The non-match scores are in red and the match scores are in blue. The dotted line indicates the highest non-match score (BKNM) of the product of the breech face score and the firing pin score. There are two data points at (27,233) which indicate very high scores for non-matching firing pin scores. In general the firing pin non-match scores seem high, indicating lower discrimination.

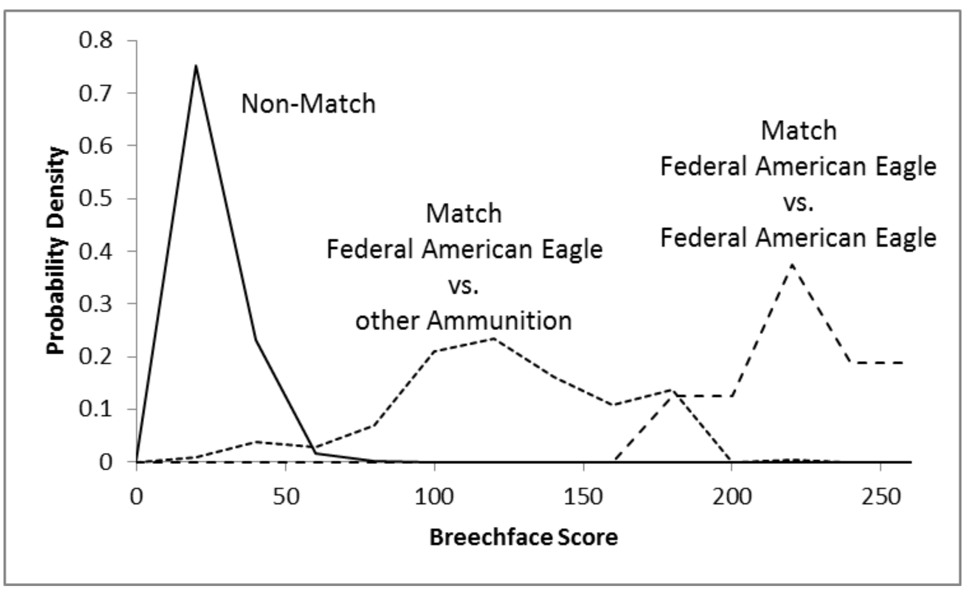

Figure 16: Probability density of breech face scores by ammunition.

In Figure 16, the breech face scores were used to compute the probability densities for the cartridge cases fired through a HiPoint C9 pistol. The solid curve is an estimate of a probability density function (pdf) for the non-match breech face scores. The short dashed line is an estimate of the pdf for breech face match scores between the input ammunition type (Federal American Eagle) matching against other ammunition types (other than Federal American Eagle). The long dashed line is an estimate of the pdf for breech face match scores of Federal American Eagle ammunition against Federal American Eagle ammunition. The separation between the two match pdfs is clear.

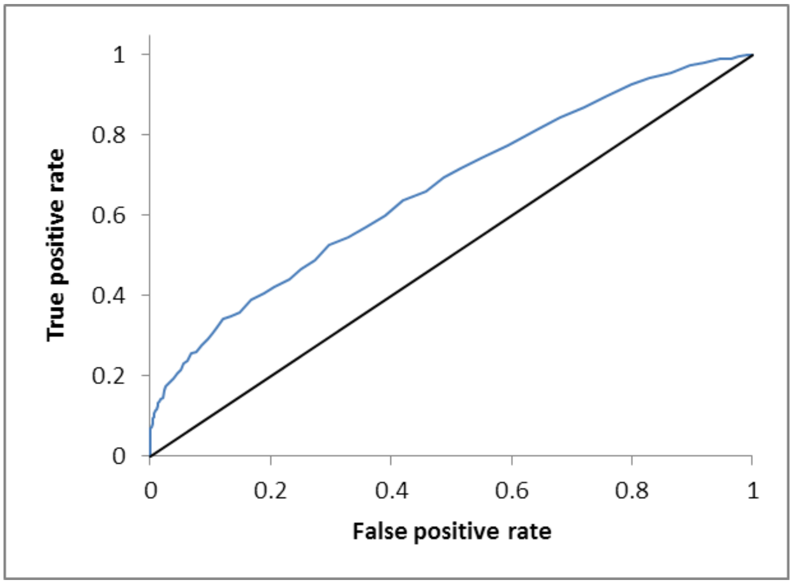

Figure 17: Example of a Receiver Operating Characteristic (ROC) curve

Figure 17 gives a ROC curve for all of the 9mm data (all firearms and all ammunition) based upon breech face scores. This ROC curve has an area under the curve (AUC) of 0.665. This implies a classification method of low to medium discriminating power. The ROC curve for the firing pin scores has an AUC of 0.786, whilst the AUC for the product of the breech face score and firing pin score is 0.810.

Conclusion

Comparison of successively fired cartridge cases suggests, from IBIS data, that the variability between shot separations is minimal. This is probably driven by the fact that the variability within shot separations is relatively large.

Sample size

Firearm examiners will usually test-fire a suspect firearm two to five times using ammunition similar to that found at the crime scene. The actual number fired is determined by laboratory policy and the experience of the firearms examiner. The examiner will then select a cartridge case from the set which is deemed to be representative of the suspect firearm. This cartridge case is then used as the known in the comparison process performed on a comparison microscope. Cartridge cases are generally entered into the IBIS system given two scenarios:

- after a comparison is made, and the suspect cartridge case is deemed not to have been fired by the suspect firearm, or
- after a single or group of cartridge cases is examined, the firearms examiner may select one or more cartridge cases for submission to IBIS.

In this project a sample of 100 cartridge cases, in most cases, were used to evaluate the variability of *same gun* (H_d) and *different gun* (H_p) scores. In order to make use of a sample of cartridge cases from a firearm to develop a *same gun* distribution, the sample distribution must be representative of the actual distribution of *same gun* scores.

Since the 9mm pistol dataset has a large number of scores, it was decided to sample distributions of the breech face and firing pin scores and to compare such sample distributions with the actual distribution for a particular model. The actual and sample distributions were compared and the variability between the two was computed using the sum of the squares. Various sample sizes were used to assess the effect of the sample size in approximating the actual distribution.

The Taurus 24/7 G2 9mm pistol was used for this test. This data set contains 951,464 records. This data set contains the IBIS scores for five pistols (X45398, X45399, X45401, X45405, and X55720).

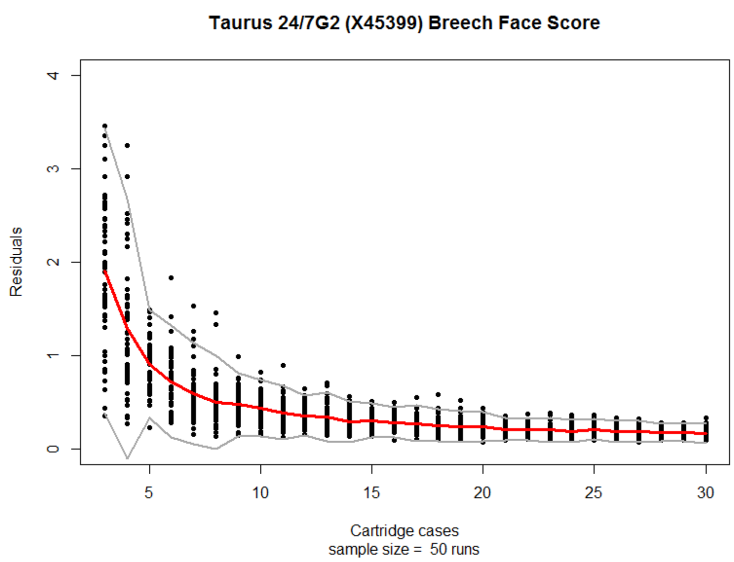

Figure 18: Sample size determination (X45399): 50 runs

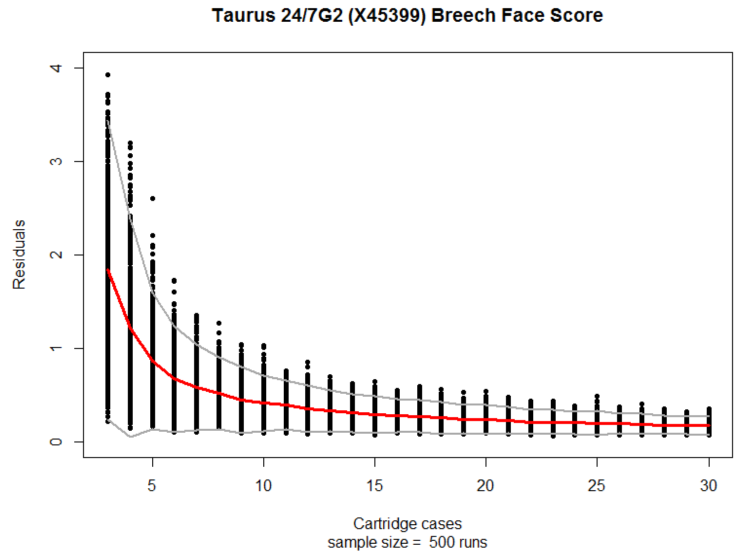

Figure 19: Sample size determination (X45399): 500 runs

Figure 20 indicates 100 simulations of the density distributions of the same gun and different gun for pistol X45399. Figure 21 indicates the actual distributions. Figure 18 and Figure 19 demonstrate the differences between the simulated and actual distributions as a function of the sample size. Figure 20 furthermore illustrates the distributions for a sample size of 10. It must be noted that 10 cartridge cases will result in 45 pairwise comparisons, and thus 45 breech face scores, to define the *same gun* distribution.

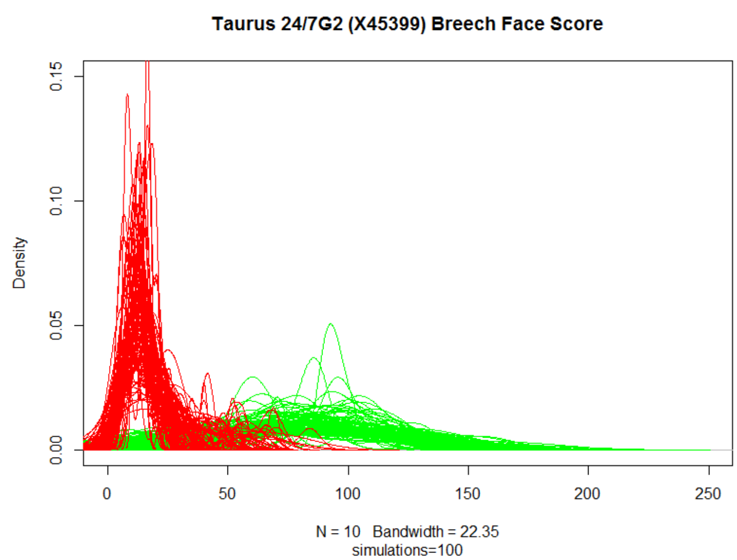

Figure 20: Simulation of distributions for BF Score (X45399)

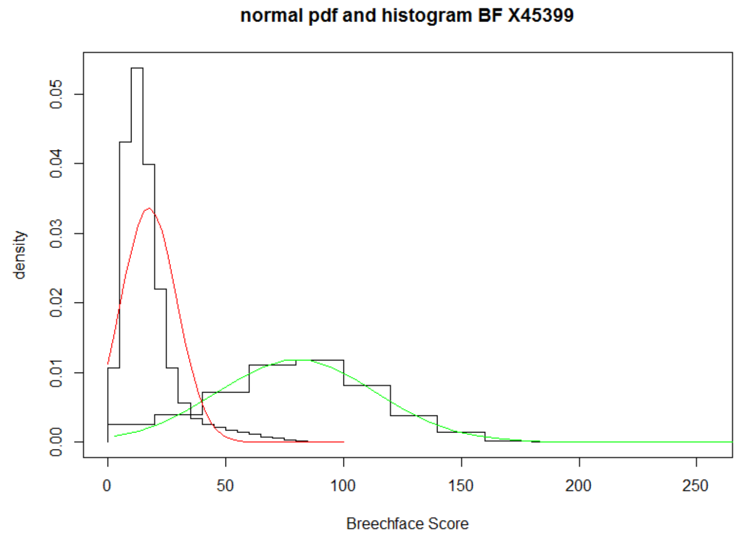

Figure 21: Actual Probability Density Function (PDF) and histogram of BF for X45399

Conclusion

In order to perform comparisons, a firearms examiner needs to produce a certain number of test fires for purposes of comparison against an unknown cartridge case (the actual number of test fires is guided through unit policies). This research examined the question of how many cartridge cases would be representative of the firearm given the observed variability in the IBIS scores. A simulation study was performed to compare the score distributions of a randomly selected sample set (*i.e.* a set of "test fires" against the distribution of a large sample or "estimated population" (generally 100 cartridge cases) of a firearm. These two distributions were compared and their similarity was measured. The larger set of "test fires," the closer the distribution of scores to that of the "population" distribution. These data suggested that a smallest sample size of test fires could be determined that would be representative of the firearm. This topic area should be researched further.

Performance of 9mm firearms

Approximately 100 cartridge cases fired by the 9mm firearms (35 pistols, 2 carbines, and 1 revolver) were submitted to IBIS and the resulting breech face and firing pin scores were analyzed using R[37] and RStudio[38]. The data were divided by model of firearm and a receiver operating characteristic (ROC) curve was computed. The area under the ROC curve was also

[37] R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

[38] RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/.

computed. All ROC calculations were performed using the ROCR package in R[39]. For each model the ROC curves and the error rate curves are given in Figure 22 through Figure 40.

Table 2 provides all of the areas under the curve (AUC) for the receiver operating characteristic (ROC) curve. These data are for the breech face scores (BF), the firing pin scores (FP), and for their product (FP*BF). The cells highlighted in green indicate which of the measures provides the most discrimination for a particular firearm. An AUC of .500 indicates that the method of classification is equal to a coin toss. A method with an AUC of 1 indicates a method that has perfect classification performance. The error rate curve illustrates how the particular cutoff (on the x-axis), in this case (BF score, FP score, and their product) affects the false positive rate (*fpr*) and the false negative rate (*fnr*). The point at which they cross is known as the equal error rate (EER). Forensic scientists would like to have a low *fpr* and thus would generally work to the right of this position with some tradeoff for the *fnr*.

[39] Sing T, Sander O, Beerenwinkel N and Lengauer T (2005). “ROCR: visualizing classifier performance in R”. Bioinformatics, 21(20), pp. 7881. http://rocr.bioinf.mpi-sb.mpg.de.

Table 2: Area under the ROC curve for all 9mm firearms[40]

Make	Model	Number of Firearms	Type	AUC_BF	AUC_FP	AUC_FPxBF
Arcus	D98	1	Pistol	0.718	0.717	0.786
Glock	19 Gen 4	1	Pistol	0.853	0.654	0.825
HiPoint	995 TS	1	Carbine	0.976	0.879	0.973
HiPoint	C9	4	Pistol	0.681	0.819	0.780
Keltec	P11	1	Pistol	0.860	0.943	0.974
Keltec	PF9	5	Pistol	0.757	0.822	0.857
Keltec	Sub2000	1	Carbine	0.621	0.977	0.899
Ruger	LC9	3	Pistol	0.737	0.811	0.835
Ruger	P95	1	Pistol	0.789	0.833	0.874
Ruger	SR9	1	Pistol	0.995	1.000	1.000
SCCY	CPX II	5	Pistol	0.546	0.602	0.574
SigSauer	P250	1	Pistol	0.998	0.995	0.999
SigSauer	SP2022	1	Pistol	0.984	0.802	0.964
Smith & Wesson	SD9VE	1	Pistol	0.850	0.837	0.883
Springfield	XD9	4	Pistol	0.656	0.770	0.768
Taurus	905	1	Revolver	0.843	0.891	0.924
Taurus	24/7 G2	5	Pistol	0.882	0.740	0.879
Taurus	Millennium Pro 111	1	Pistol	0.996	0.993	0.999
All	**All**	**38**		**0.741**	**0.756**	**0.799**

From Table 2 it can be seen that IBIS scores for the SCCY CPX II pistols performed very badly as a classifier for the same gun/different gun scenario. The best performers were for the Ruger SR9, SigSauer P250, and the Taurus Millennium Pro 111. Of the 18 models represented in Table 2, four had breech face score as the best performer, four with the firing pin score, and 10 with the product of both. This is illustrative that in many cases both scores should be considered. This is not easily achieved with the current configuration of the IBIS. Overall, the product of the scores is the best performer in classification.

[40] Best performing classifiers colored greem.

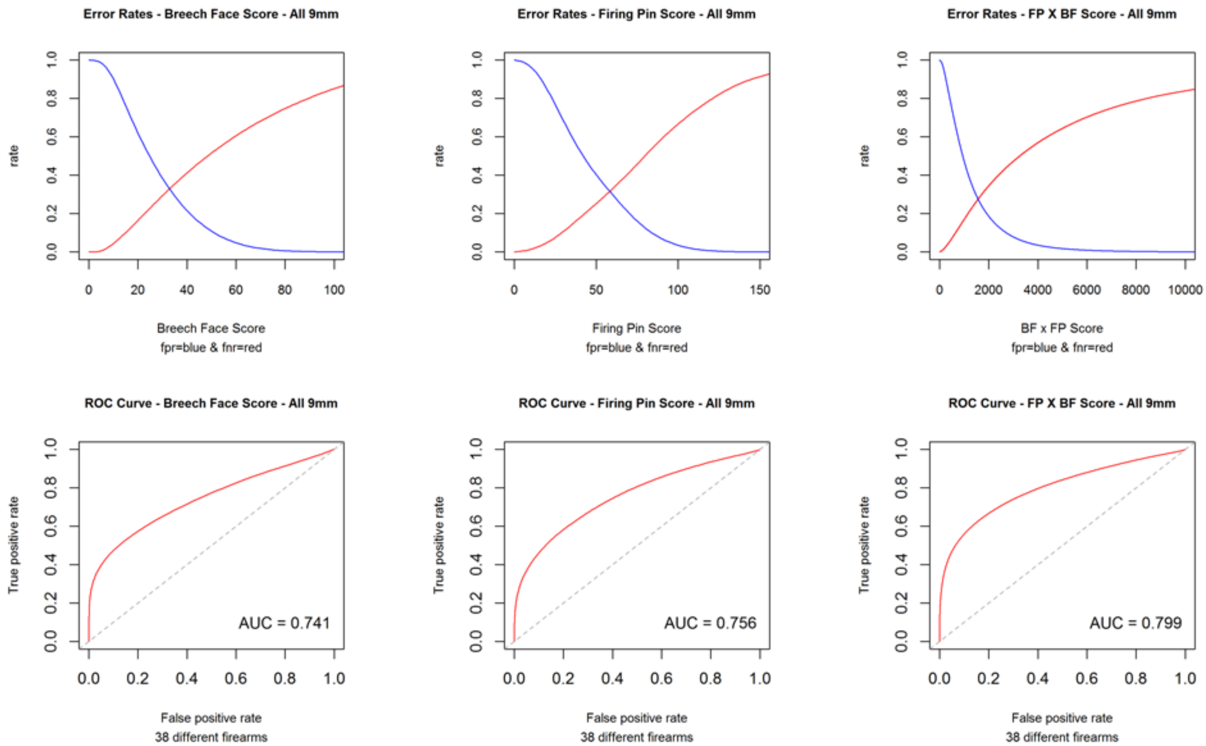

Figure 22: ROC and error rate curves for all 9 mm firearms

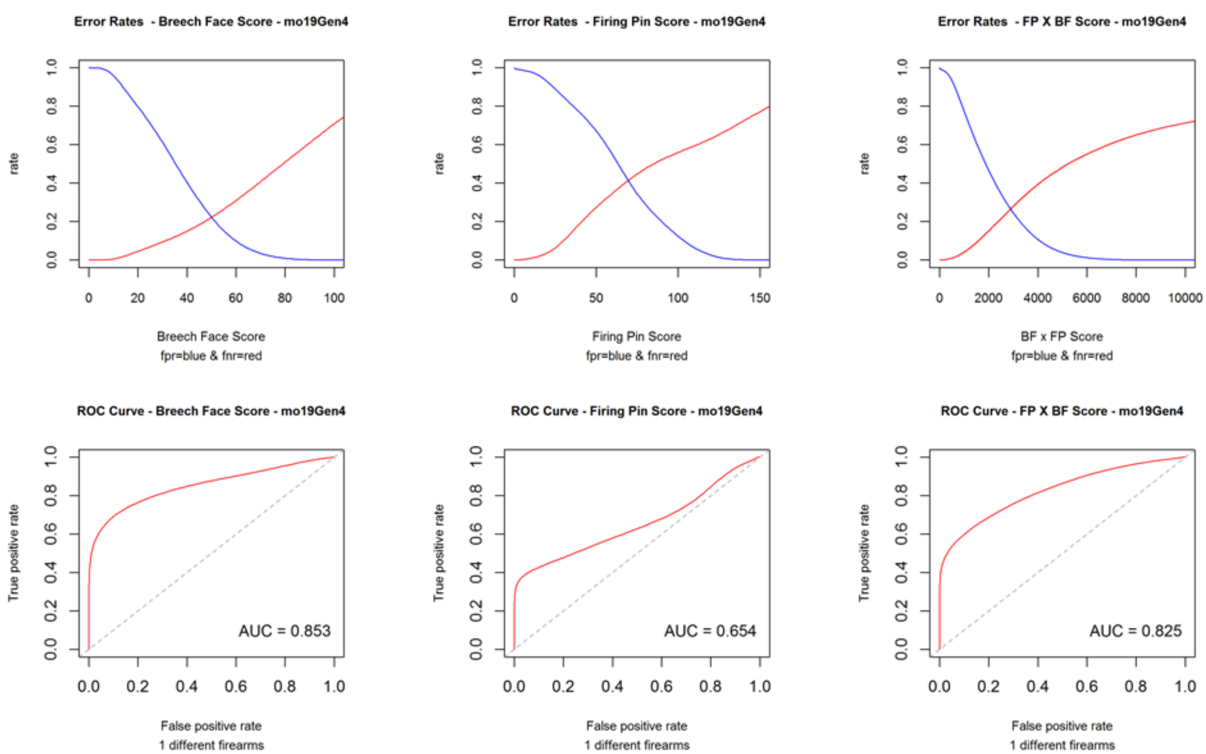

Figure 23: ROC and error rate curves for all Glock 19 Gen 4 (9 mm) firearms

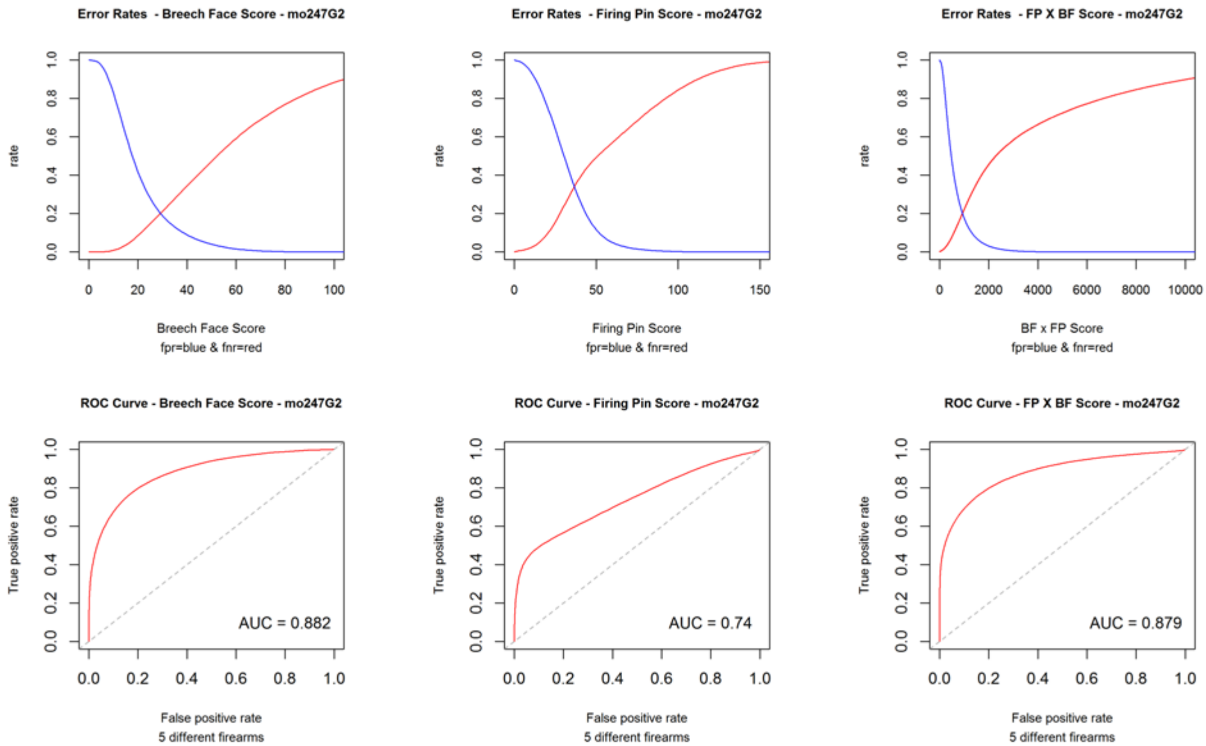

Figure 24: ROC and error rate curves for all Taurus 24/7 G2 (9 mm) firearms

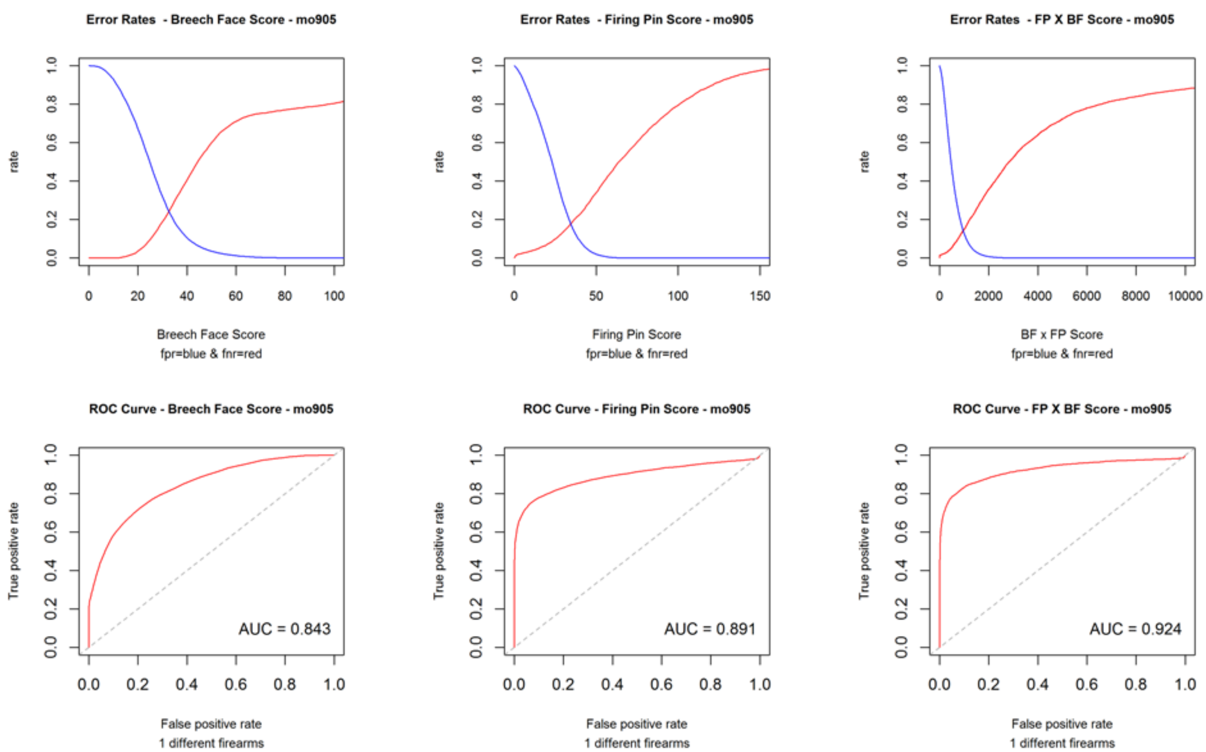

Figure 25: ROC and error rate curves for all Taurus Model 905 (9 mm) firearms

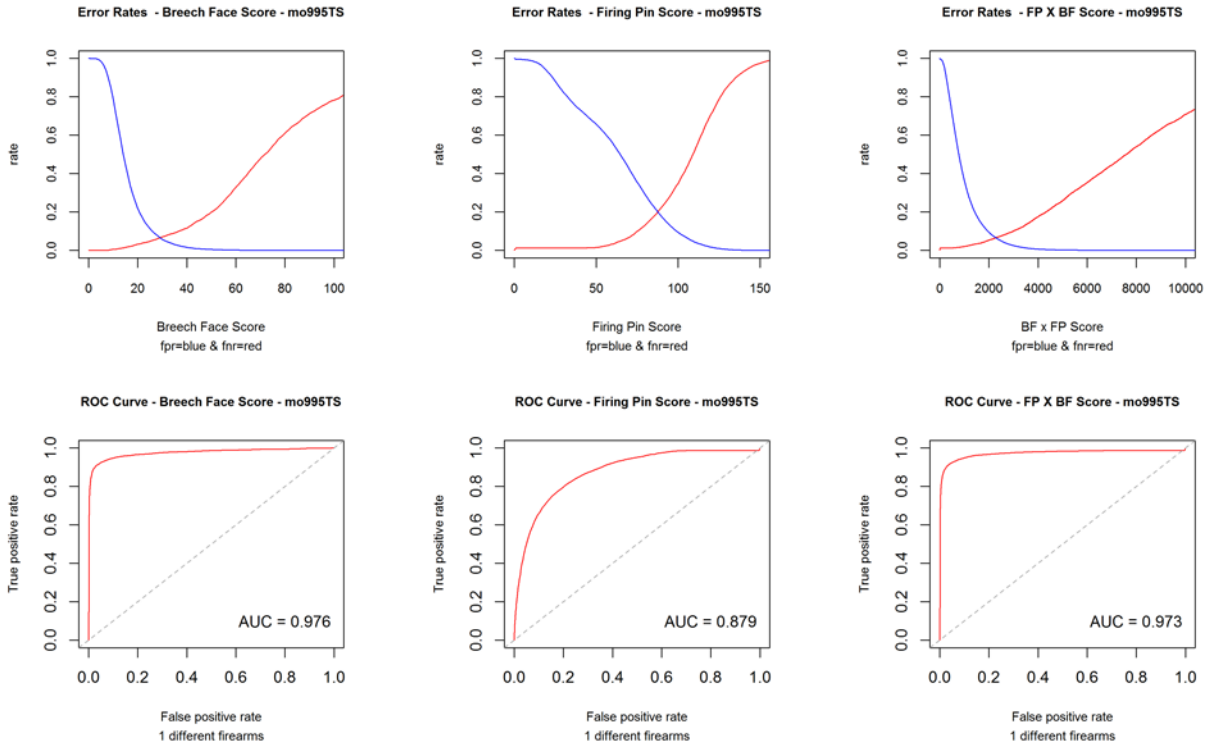

Figure 26: ROC and error rate curves for all HiPoint 995 TS (9 mm) firearms

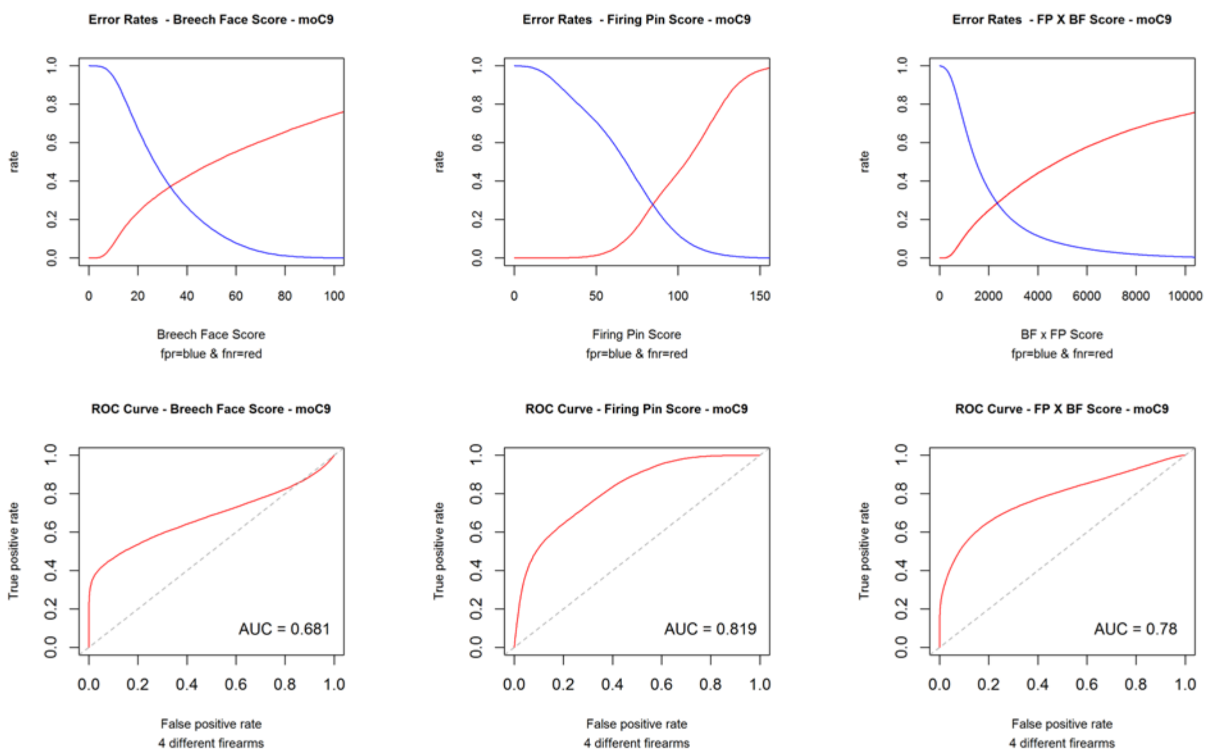

Figure 27: ROC and error rate curves for all HiPoint C9 (9 mm) firearms

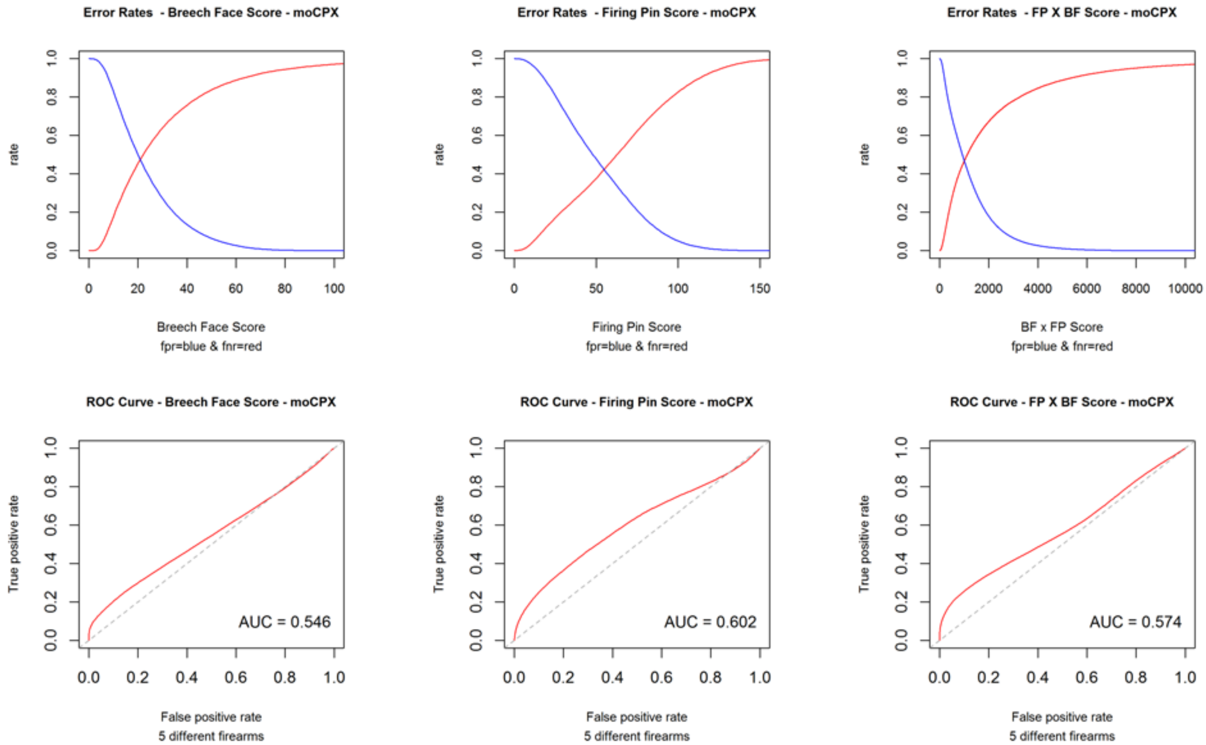

Figure 28: ROC and error rate curves for all SCCY CPX II (9 mm) firearms

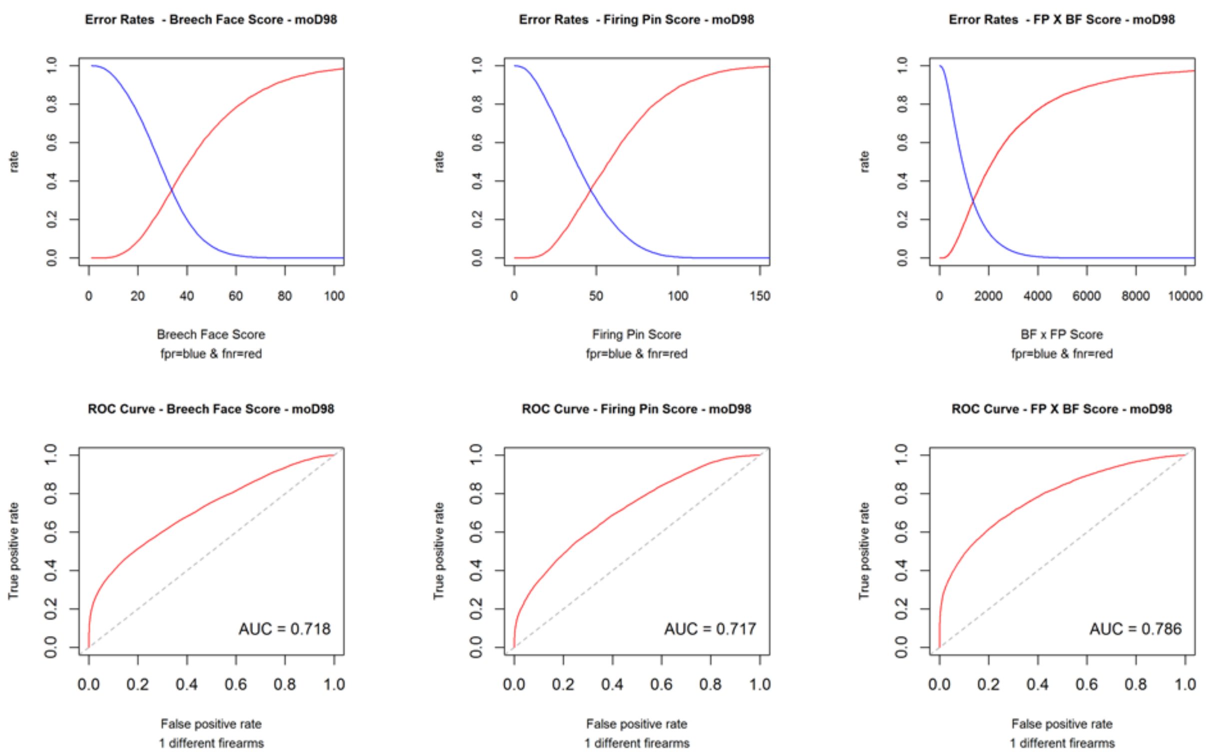

Figure 29: ROC and error rate curves for all Arcus D98 (9 mm) firearms

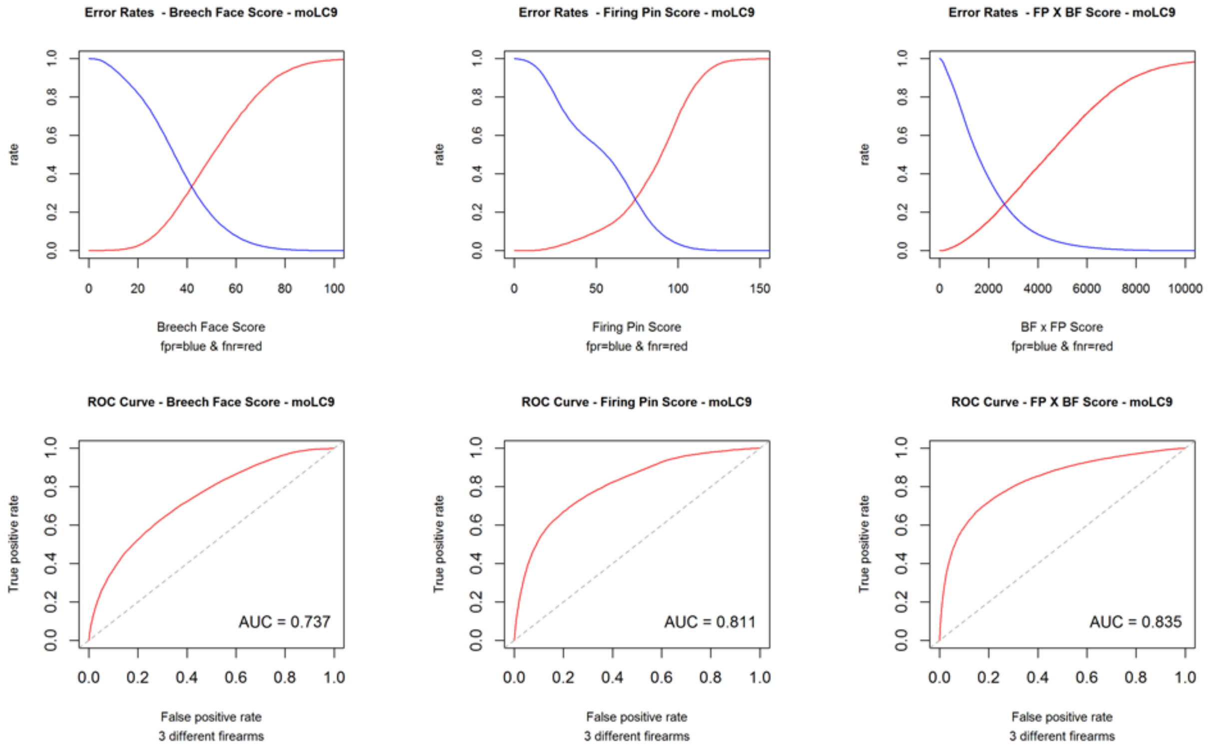

Figure 30: ROC and error rate curves for all Ruger LC9 (9 mm) firearms

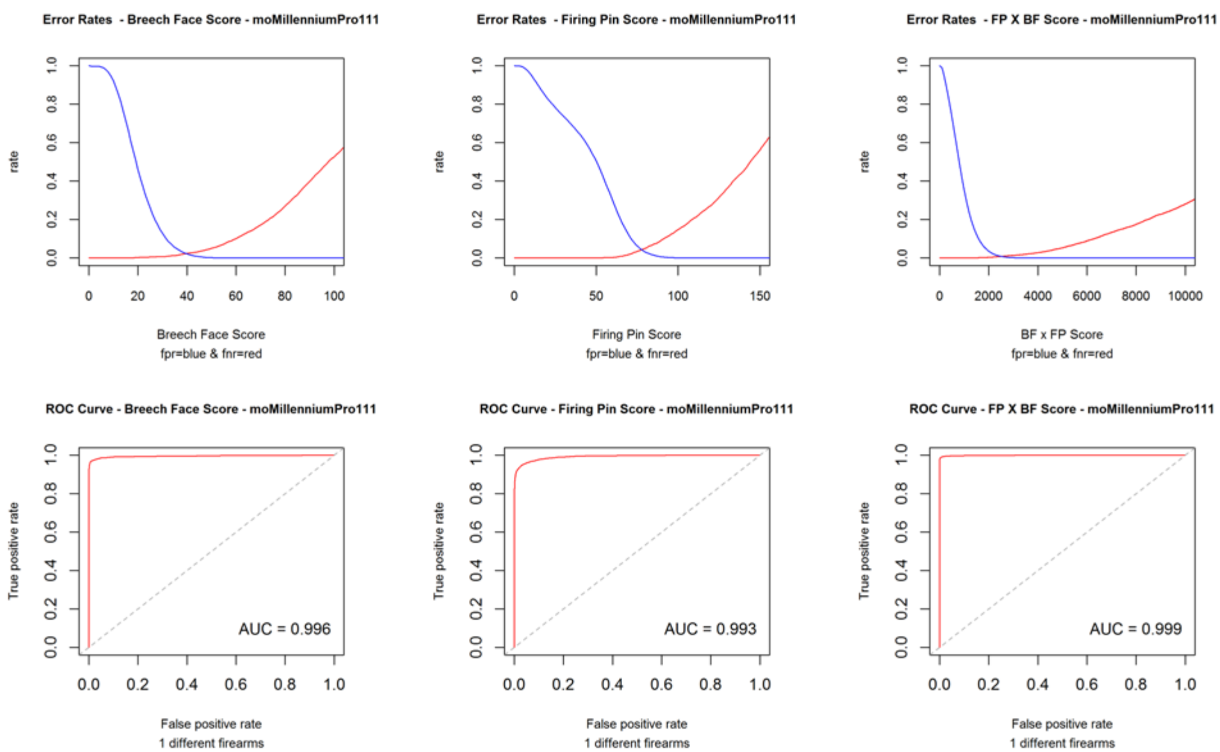

Figure 31: ROC and error rate curves for all Taurus Millennium Pro 111(9 mm) firearms

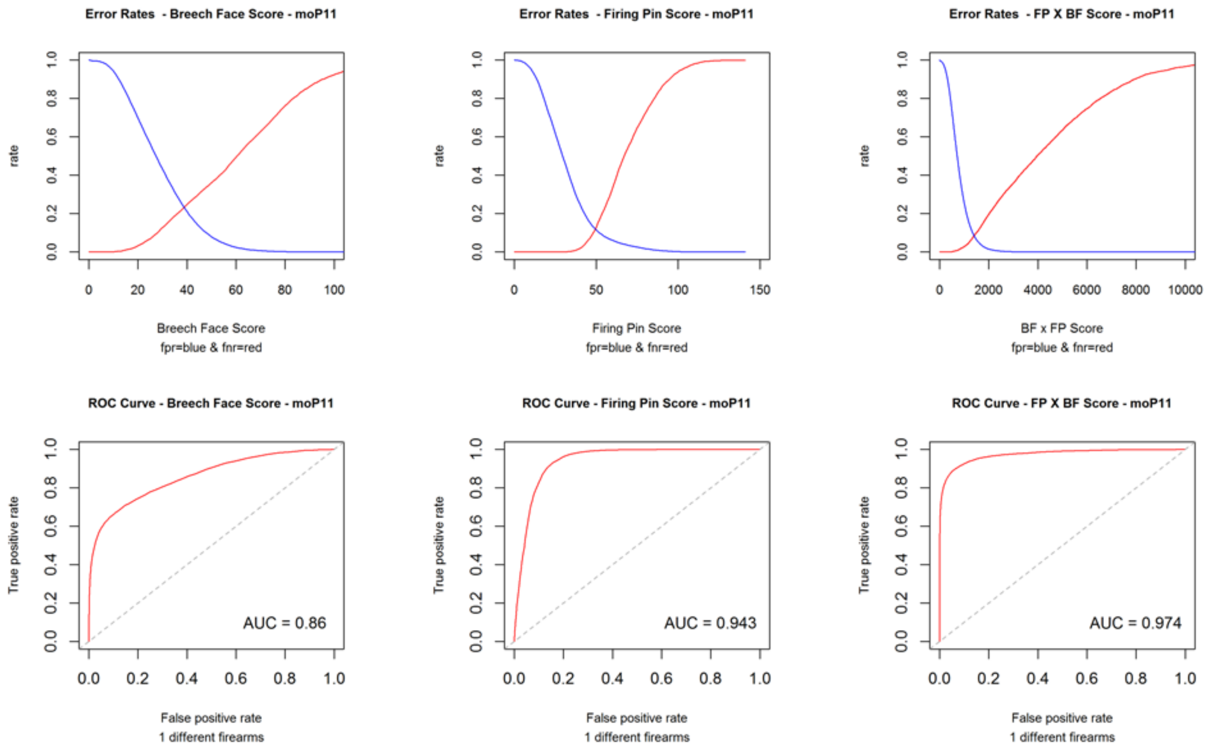

Figure 32: ROC and error rate curves for all Keltec P11 (9 mm) firearms

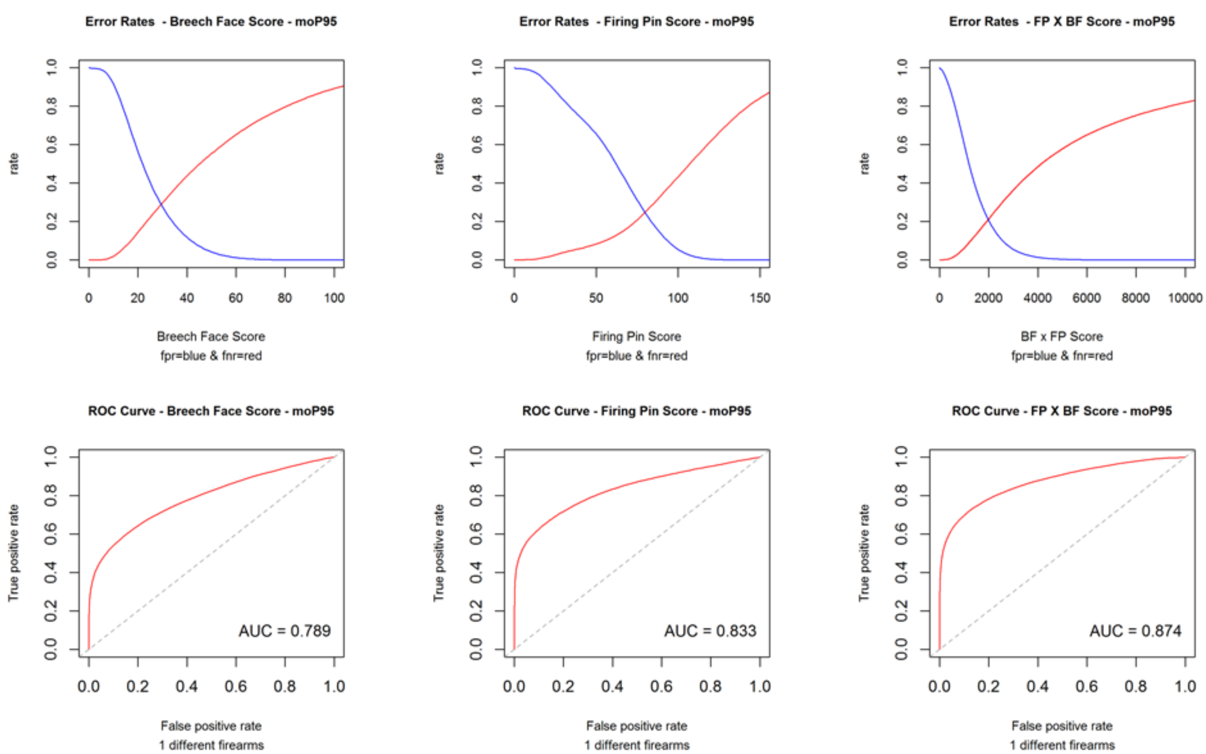

Figure 33: ROC and error rate curves for all Ruger P95 (9 mm) firearms

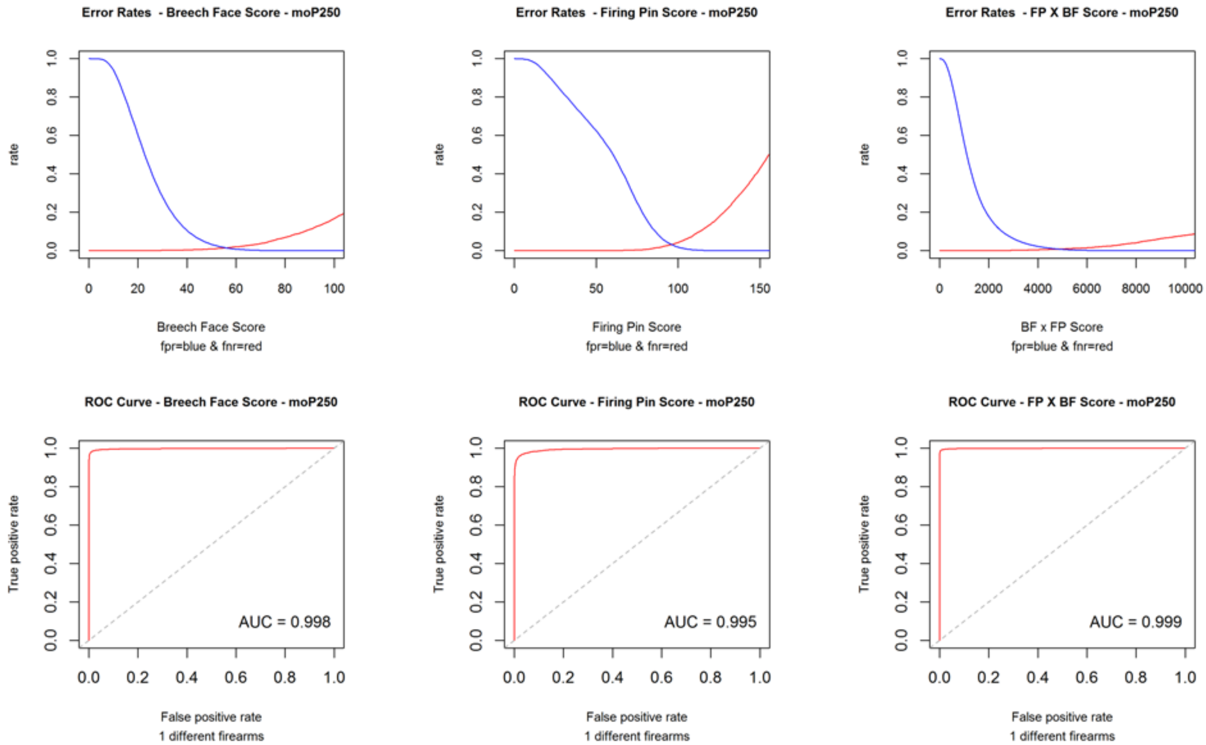

Figure 34: ROC and error rate curves for all SigSauer P250 (9 mm) firearms

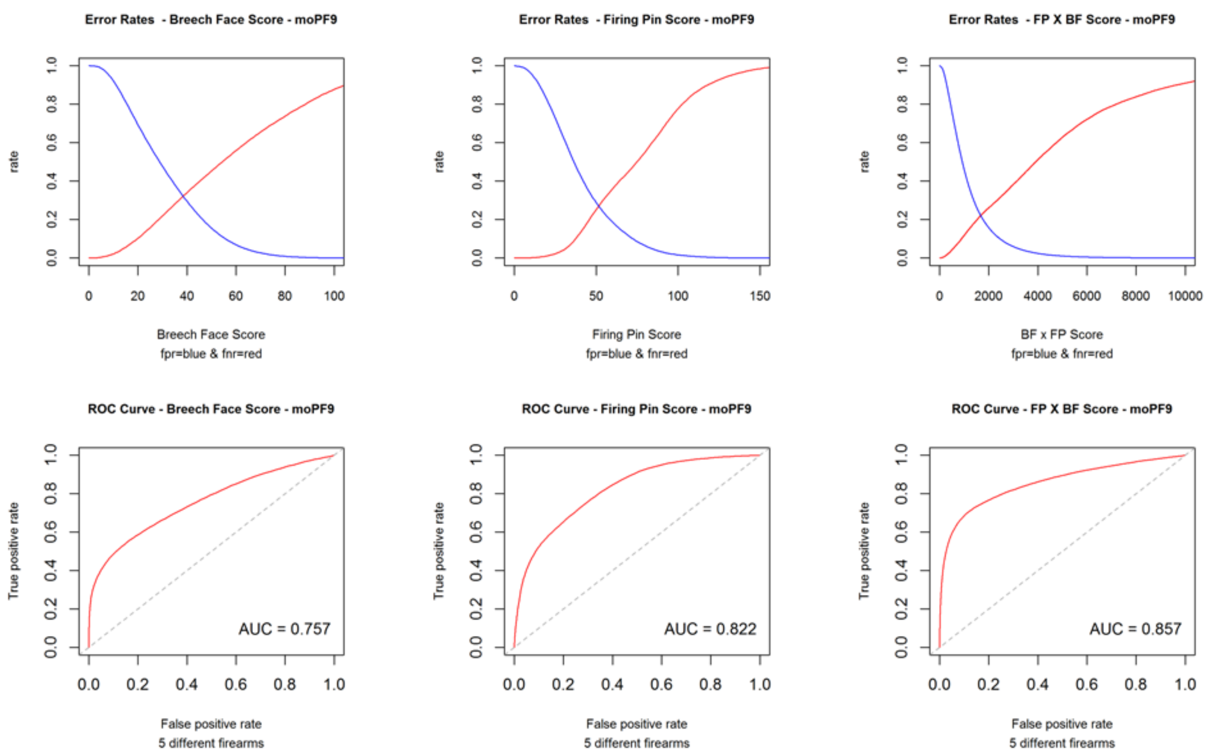

Figure 35: ROC and error rate curves for all Keltec PF9 (9 mm) firearms

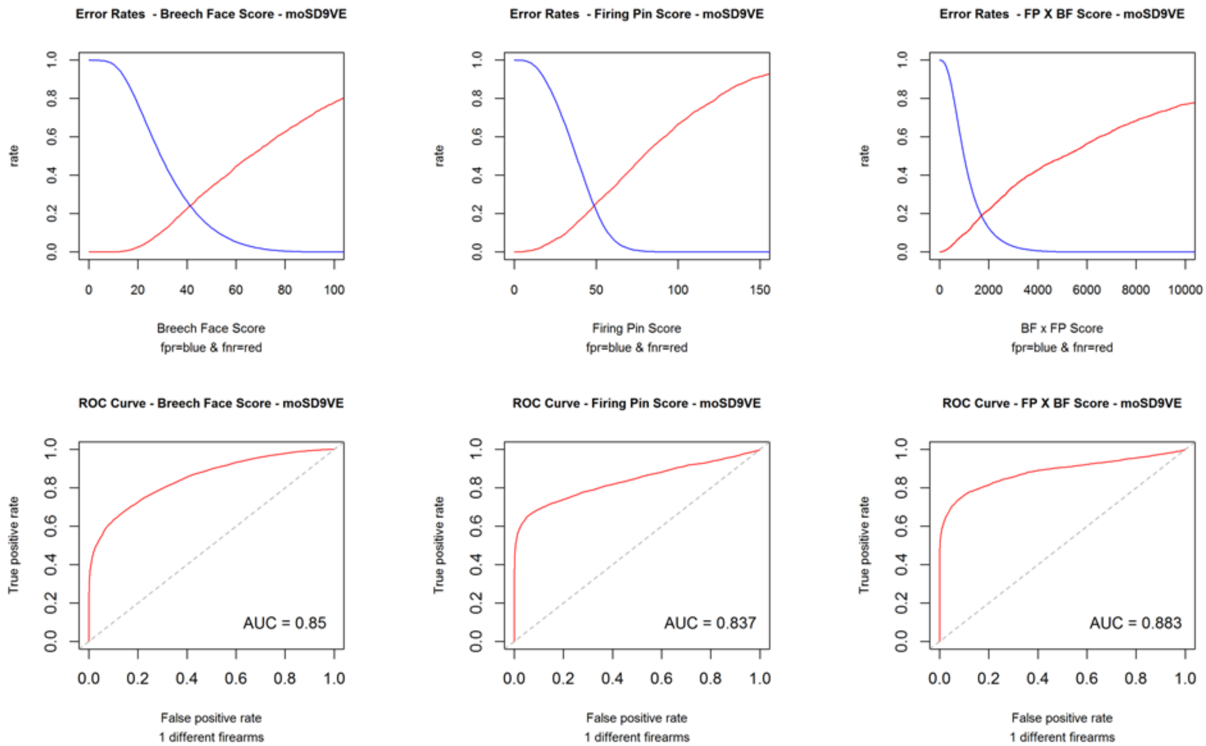

Figure 36: ROC and error rate curves for all Smith & Wesson SD9VE (9 mm) firearms

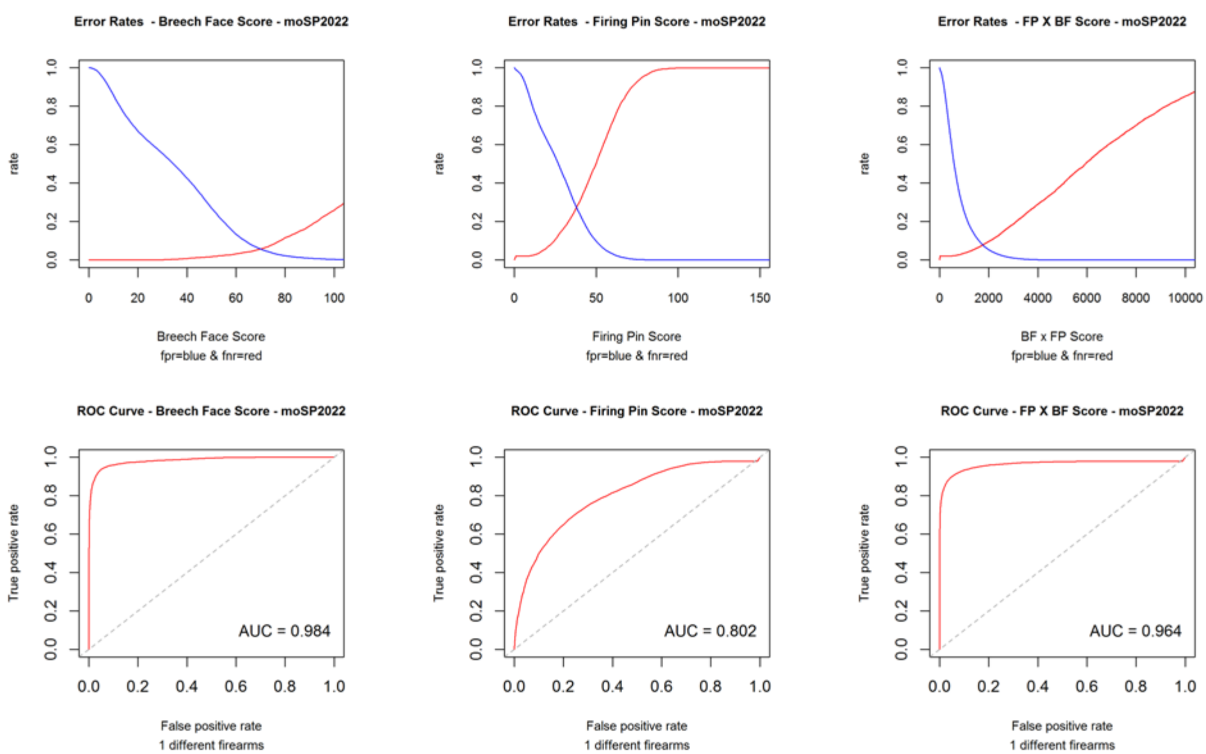

Figure 37: ROC and error rate curves for all SigSauer SP2022 (9 mm) firearms

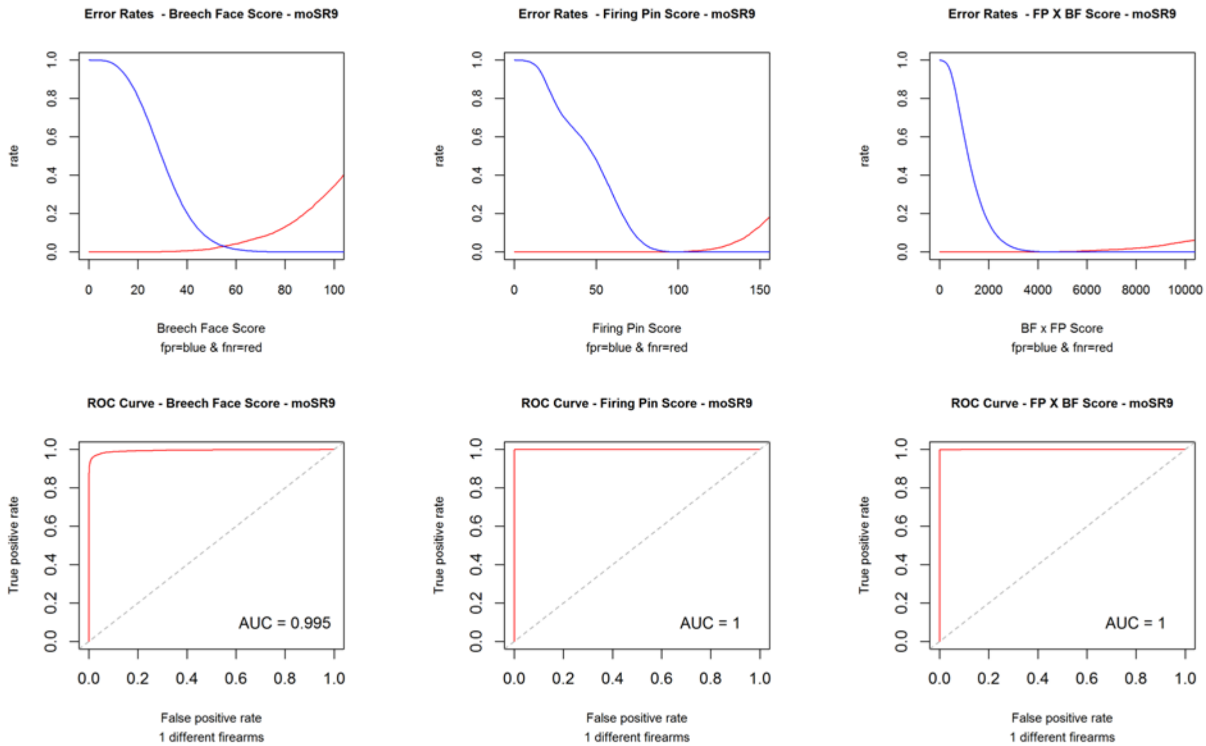

Figure 38: ROC and error rate curves for all Ruger SR9 (9 mm) firearms

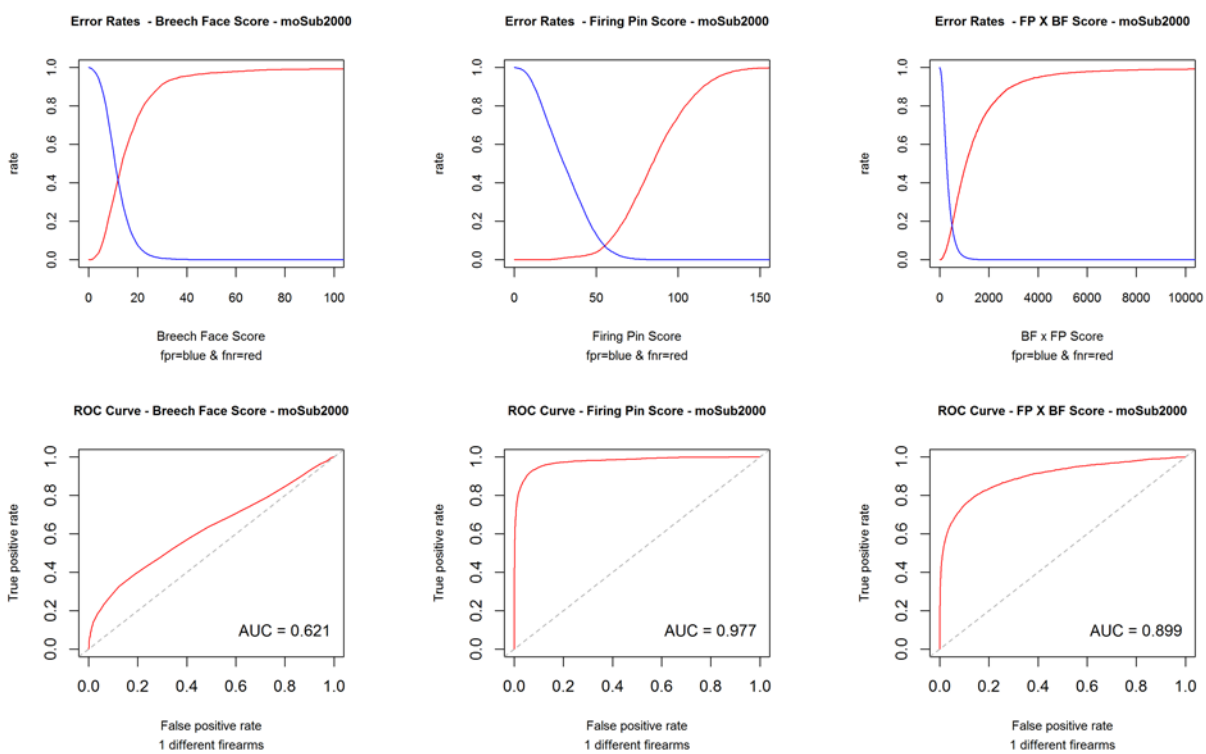

Figure 39: ROC and error rate curves for all Keltec Sub2000 (9 mm) firearms

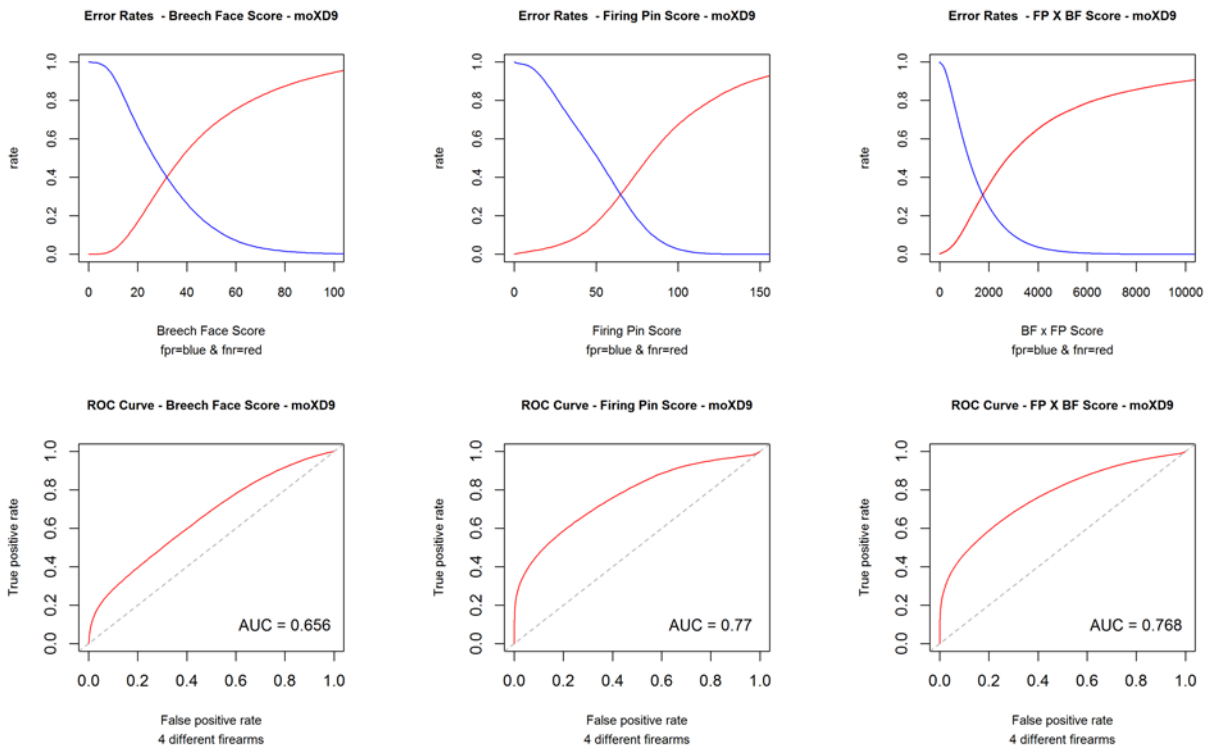

Figure 40: ROC and error rate curves for all Springfield XD9 (9 mm) firearms

Conclusion

The breech face (BF) and firing pin (FP) scores, as well as their product (BFxFP), generated by the IBIS were used to assess the ability of the system to classify an "unknown" cartridge case into a same-gun or different-gun category. There were 38 9mm Luger firearms (represented by 10 manufacturers and 18 models) used in this study. For the Ruger SR9, both the FP score and the BFxFP score were perfect classifiers. The BF score was the best classifier for four models (Glock 19, HiPoint 995TS, SigSauer SP2022, and the Taurus 24/7 G2), the FP score was the best classifier for five models (HiPoint C9, Keltec Sub 2000, Ruger SR9, SCCY CPX II, and the Springfield XD9), and the BFxFP score was the best classifier for nine models (Arcus D98, Keltec P11, Keltec PF9, Ruger LC9, Ruger P95, Ruger SR9, SigSauer P250, S&W SD9-VE, Taurus 905, and the Taurus Millennium Pro 111). The IBIS system does not provide for an easy means to use the combination of the BF and FP scores. The ability to order candidate lists through the combination of scores will be of value to firearms examiners (especially so in the 3D system). Since the markings that appear on the breech face and firing pins (or strikers) are made through independent manufacturing operations, the score generated through the IBIS comparisons are also independent. Generally, all of the classifiers performed well but the SCCY CPX II pistols were the worst in all three measures. This was due to markings that were difficult for IBIS to interpret, but would be easy identifications for a firearms examiner.

System performance studies

Four sets of samples were received from US Army Criminal Investigation Laboratory (USACIL), a division of the Defense Forensic Science Laboratory (DFSC). These samples used to assess the repeatability, reproducibility, comparison, and "blind" samples.

The following methods were used to study the samples:

> Repeatability: The single cartridge case was mounted in the IBIS system. Five IBIS case files were created with their respective exhibits. For each case file the analyst imaged the cartridge case without removing or readjusting the position of the cartridge case between captures. Each analyst repeated this procedure.
>
> Reproducibility: The single cartridge case was mounted in the IBIS system. Five IBIS case files were created with their respective exhibits. The cartridge case was dismounted, remounted, and the position readjusted between captures. Each analyst repeated this procedure.
>
> Comparison: For each comparison set, the three known samples (K1, K2, and K3) and the one questioned sample (Q1) were captured by the IBIS system. Each analyst repeated this procedure.
>
> Blind: Five blind samples (1 – 5) were captured by the IBIS system. Each analyst repeated this procedure.

After all captures were completed, all of the exhibits were re-correlated and the data processed according to the standard procedure.

Data analysis of IBIS breech face and firing pin scores

In order to define a methodology for the analysis of the data the following standard definition (where available) were used to develop a useable definition. According to the EURACHEM Guide[41], repeatability may be defined as:

> "Precision under repeatability conditions, i.e. conditions where independent test results are obtained with the same method on identical test items in the same laboratory by the same operator using the same equipment within short intervals of time."

[41] Eurachem Guide: The Fitness for Purpose of Analytical Methods, December 1998, http://www.eurachem.org/images/stories/Guides/pdf/valid.pdf

Furthermore, it states that the repeatability standard deviation is the "standard deviation of test results obtained under repeatability conditions." In addition, it notes that the repeatability standard deviation is a "measure of dispersion of the distribution of test results under repeatability conditions. Similarly 'repeatability variance' and 'repeatability coefficient of variation' could be defined and used as measures of the dispersion of test results under repeatability conditions."

For this study the definition of repeatability that is utilized is a variation on the definition as proposed by Eurachem. The measure of precision used to evaluate the repeatability is the coefficient of variation for both breech face and firing pin scores.

Precision under repeatability conditions, i.e. conditions where independent test results are obtained with the same acquisition method on repetitive imaging of an identical test item in the same laboratory by the same operator using the same equipment within short intervals of time.

Also from the EURACHEM Guide, reproducibility may be defined as:

> "Precision under reproducibility conditions, i.e. conditions where test results are obtained with the same method on identical test items in different laboratories with different operators using different equipment."

It also notes that a "valid statement of reproducibility requires specification of the conditions changed. Reproducibility may be expressed quantitatively in terms of the dispersion of the results."

The reproducibility standard deviation is defined as the "standard deviation of test results obtained under reproducibility conditions." Similar to repeatability, this is a "measure of dispersion of the distribution of test results under reproducibility conditions. Similarly 'reproducibility variance' and 'reproducibility coefficient of variation' could be defined and used as measures of the dispersion of test results under reproducibility conditions."

For this study the definition of reproducibility is utilized is a variation on the definition as proposed by Eurachem. The measure of precision used to evaluate the reproducibility is the coefficient of variation for both breech face and firing pin scores. Precision under reproducibility conditions, *i.e.* conditions where test results are obtained with the same acquisition method on separate imaging of an identical test item in the same laboratory with the same operators using the same equipment.

Repeatability

The data were sliced to obtain the breech face (BF) and firing pin scores (FP) of each analyst against themselves as per the adopted definition of repeatability. Figure 41 provides a plot of the coefficient of variation[42] (CoV) for both the BF and FP scores. Since the CoV takes both the mean (μ) and standard deviation (σ) into account, this implies that an analyst with a low μ and a low σ could have the same CoV as an analyst with a higher μ and a higher σ. The maximum CoV BF is less than 11%, whilst the maximum CoV for FP is less than 30%. This variability between examiners may seem high for the FP CoV, but it must be remembered that the score values obtained in this study are extremely high scores, not usually seen in casework. It may also amplify the fact that small changes in light may have a considerable impact on the net score.

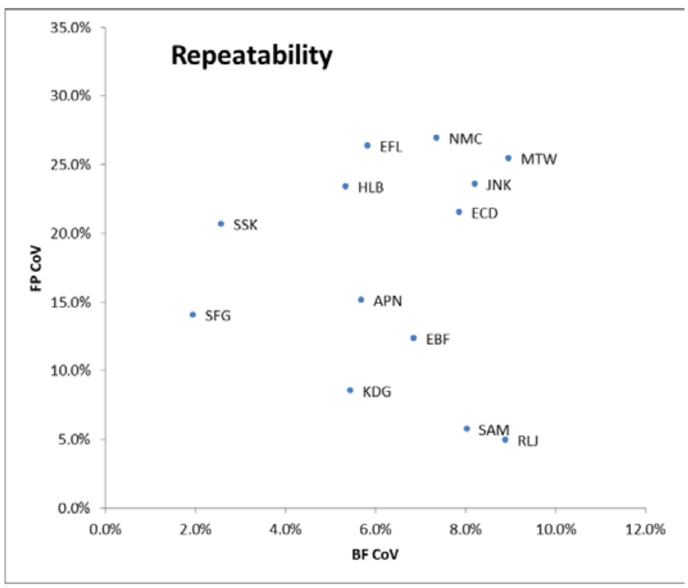

Figure 41: Repeatability – firing pin coefficient of variation versus breech face coefficient of variation

The repeatability data for the breech face scores are given in Figure 42. The letters given on the top of the figure indicate the groups of which each analyst is a member. A group is created when the means of the members are not significantly different. Each group is assigned a letter name. For example, ECD, RLJ, and SAM are all members of group 'c', whilst HLB is a member of groups 'a' and 'b'. The means of the groups with the same letter are not significantly different. In this instance there are three significantly different groups of means for BF. The firing pin repeatability is given in Figure 43. For firing pin scores there are seven significantly different groups of means.

[42] The coefficient of variation is calculated by using all of the relevant scores per analyst to determine the mean (μ) and the standard deviation (σ) of the scores. $CoV = \frac{\mu}{\sigma} \times 100\%$

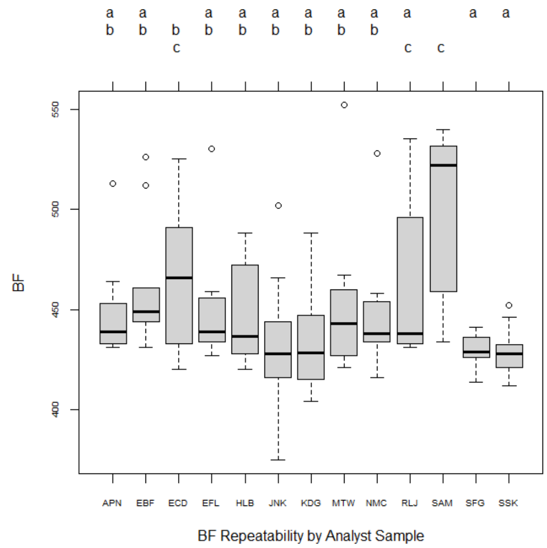

Figure 42: Breech face repeatability

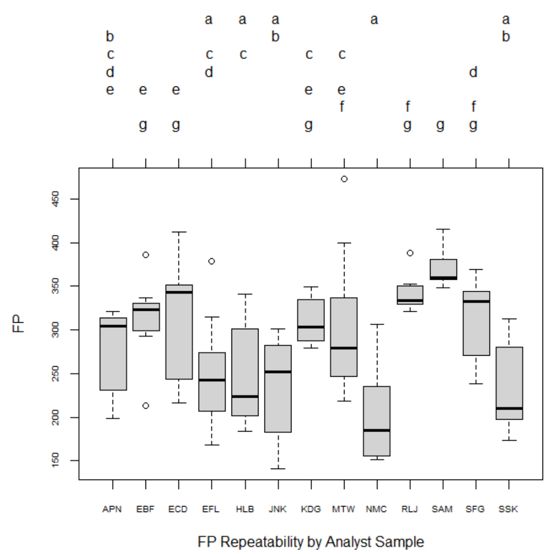

Figure 43: Firing pin repeatability

Reproducibility

Similar to the repeatability study, the data were sliced to obtain the breech face (BF) and firing pin scores (FP) of each analyst against themselves as per the adopted definition of reproducibility. Figure 44 provides a plot of the coefficient of variation for both the BF and FP Scores. The maximum CoV BF is slightly less than 12% (similar to the repeatability value), whilst the maximum CoV for FP is less than 30%. Apart from one examiner (NMC, an inexperienced operator), the rest of the CoV's are more clustered.

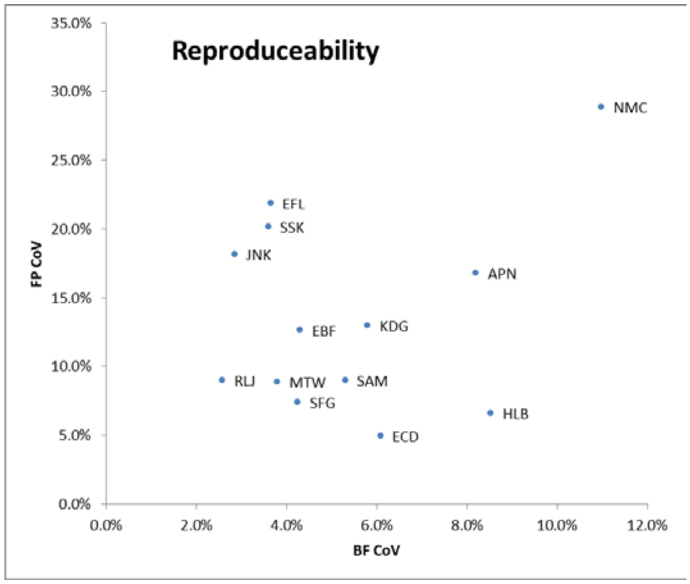

Figure 44: Reproducibility - FP CoV and BF CoV

The reproducibility data for the breech face scores are given in Figure 45. For the BF reproducibility, there are four significantly different groups of means for BF. The firing pin score reproducibility is given in Figure 46. For firing pin scores there are six significantly different groups of means.

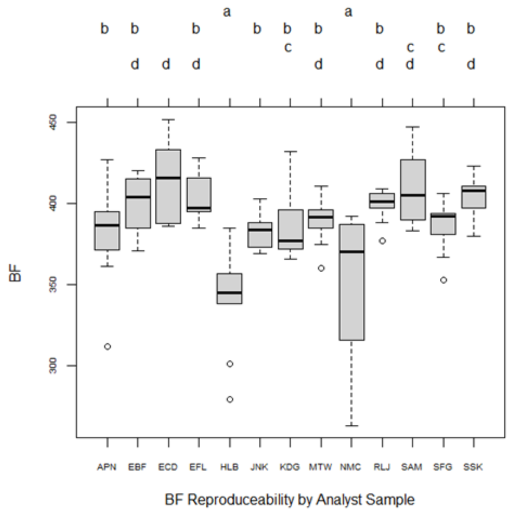

Figure 45: BF Reproducibility

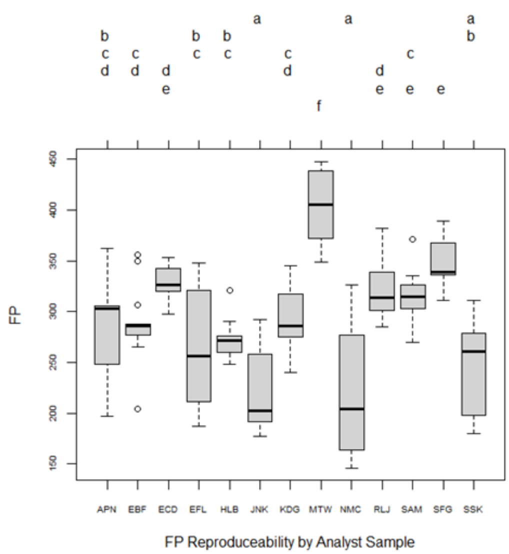

Figure 46: FP Reproducibility

Another view of repeatability and reproducibility is given in Figure 47 (breech face) and Figure 48 (firing pin). The plots provide these two metrics for each score. The line in the graph indicates where the CoV for each is equal. CoV's above the line indicate better CoV in repeatability then in reproducibility. Analysts closer to the origin have better overall repeatability and reproducibility for FP and BF scores.

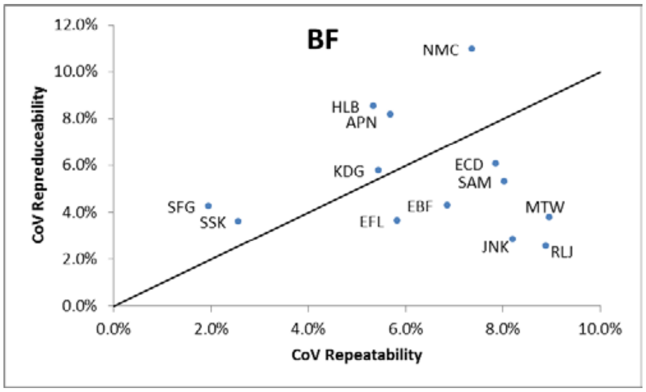

Figure 47: Repeatability and reproducibility of the IBIS breech face score

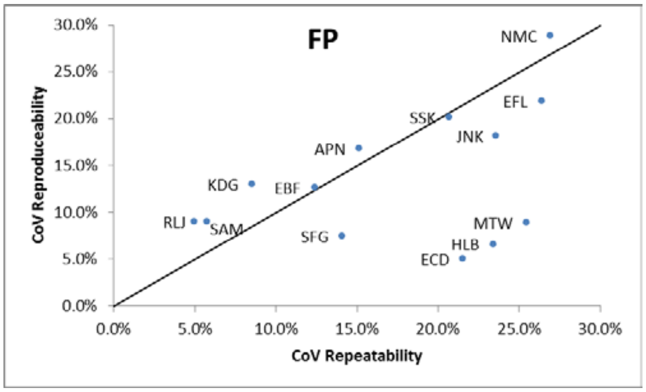

Figure 48: Repeatability and reproducibility of IBIS firing pin scores

All reproducibility studies were conducted using a cartridge case fired by a Ruger P95DC.

Blind studies

For the blind and comparison studies, the preliminary Bayesian network (BN) in Figure 49 was used to compute the likelihood ratios**.**

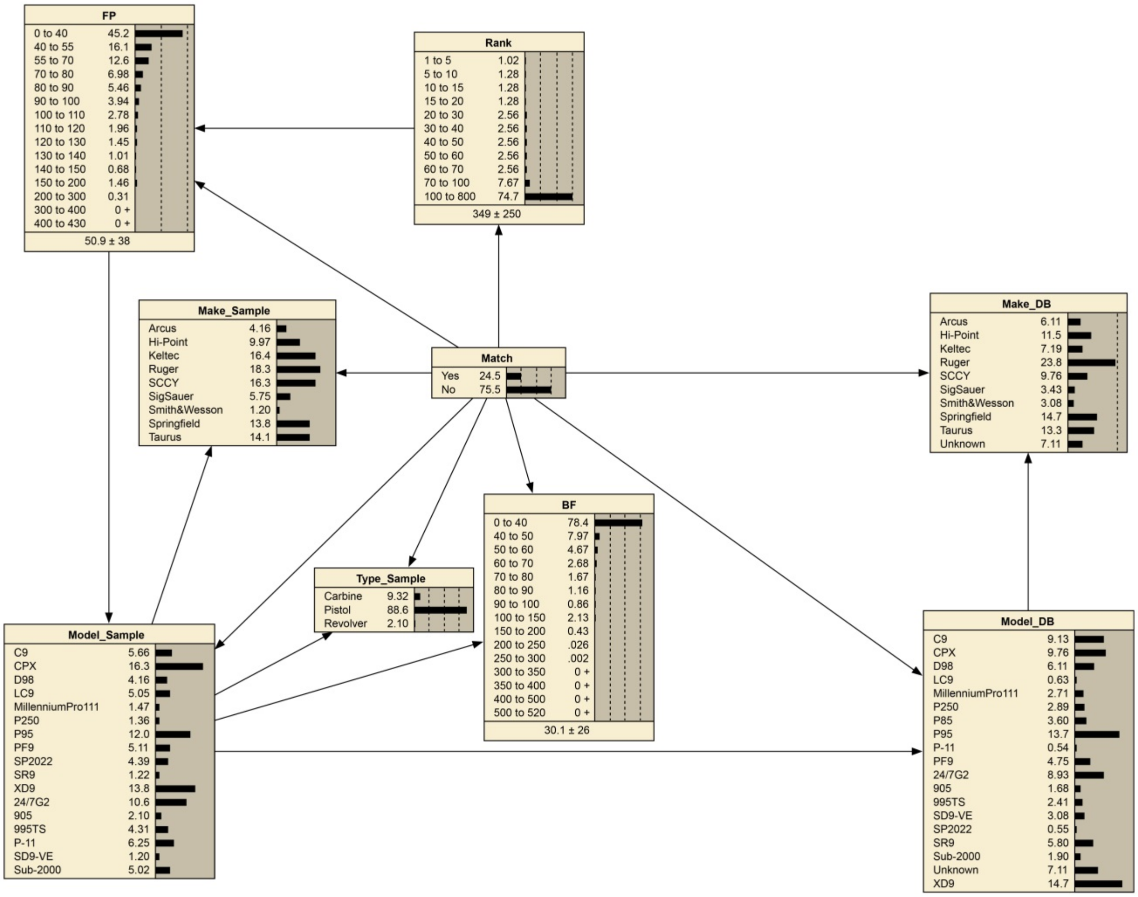

Figure 49: 9mm Bayesian Network

Blind #1:

In each case, the evaluation for the determination of the model of the firearm a new LR had to be computed. The Prior Odds for each model were computed as follows:

For each model, P(model)[43] was taken directly from the BN, and

[43] probability for a particular model

Equation 4: Calculation of prior odds for firearm model

$$P(\neg model) = \sum_{i=1}^{Model\_Sample \neq model} P(i) = 1 - P(model)$$

Therefore,

$$prior\ odds = \frac{P(model)}{P(\neg model)}$$

A similar approach is used for the calculation of the posterior odds after the instantiation of the various nodes. The $\log_{10}$(likelihood ratio) (LLR) is provided to allow for direct contrast of H_d and H_p (same magnitudes but opposite directions).

For each of the 5 blind samples, 4 sets of data (lowest numerical rank, highest BF, highest FP, and highest FPxBF) were used to select the data. Once the data was sorted a full set of Rank, FP and BF were entered into the BN (Figure 49).

All of the results are given in Figure 50 through Figure 69. All of the plots use the same x-axis scales to allow for comparison between the various scenarios. Consider Figure 50. Results which are to the right of the y-axis support the selection of a particular model. Results to the left of the y-axis provide support for the particular model not being the one which fired the question cartridge case. The extent to which the results deviate from the y-axis demonstrate the magnitude of agreement with the proposition. It can be seen that there is very strong[44] evidence to support the proposition that the cartridge case was not fired from the Ruger SR9, the HiPoint 995TS, or the Taurus Millennium Pro 111. There is moderate evidence to support the proposition that the cartridge case was fired from a Smith & Wesson SD9-VE. There is limited evidence to support the proposition that the cartridge case was fired from a Keltec PF9 and that it was not fired by a Keltec P11.

[44] Evett, I. W., G. Jackson, J. A. Lambert, and S. McCrossan. 2000. The impact of the principles of evidence interpretation on the structure and content of statements. Science & Justice 40 (4): 233–9

LLR	verbal convention
0-1	limited evidence to support
1-2	moderate evidence to support
2-3	moderately strong evidence to support
3-4	strong evidence to support
>4	very strong evidence to support

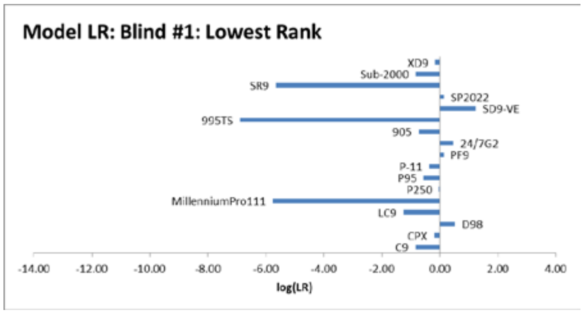

Figure 50: Blind #1 - Lowest rank

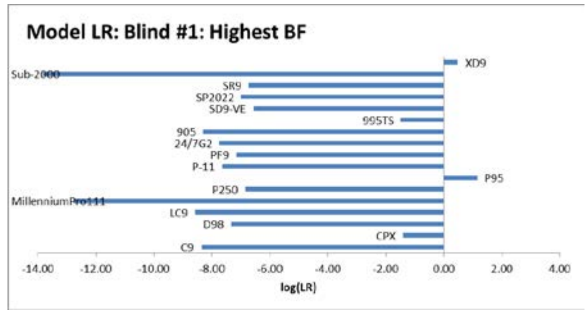

Figure 51: Blind #1 - Highest BF

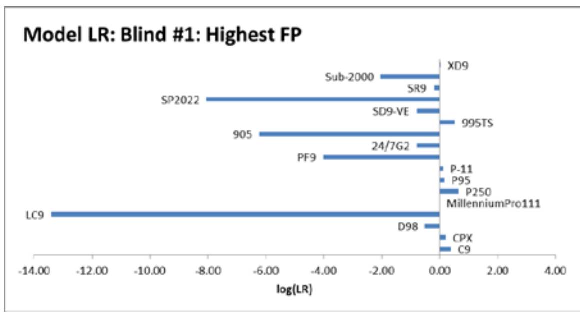

Figure 52: Blind #1 - Highest FP

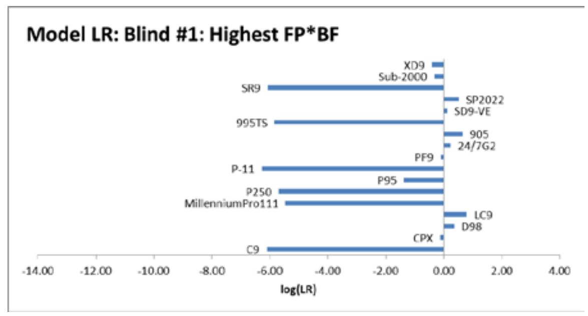

*Figure 53: Blind #1 - Highest FP*BF*

The ground truth for Blind #1 was Ruger P95DC.

Blind #2:

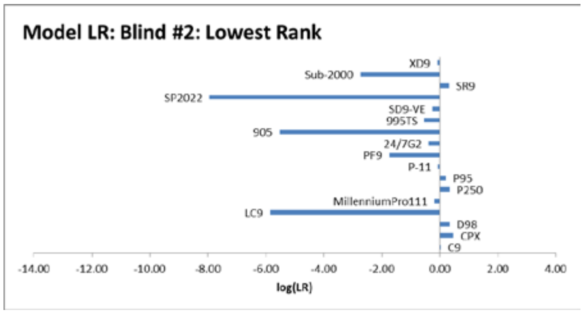

Figure 54: Blind #2 - Lowest Rank

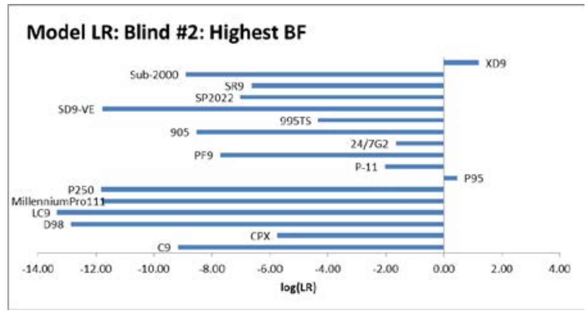

Figure 55: Blind #2 - Highest BF

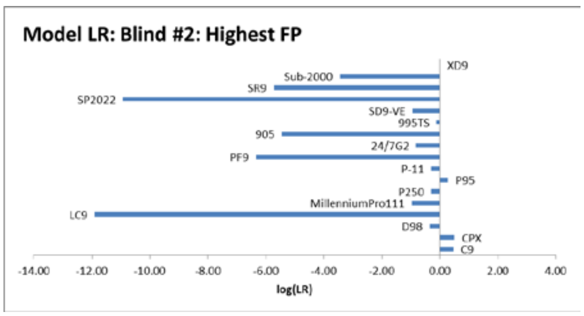

Figure 56: Blind #2 - Highest FP

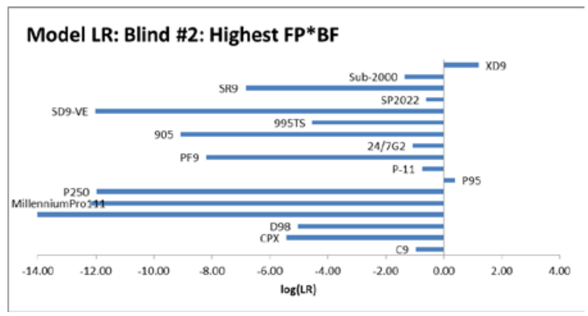

*Figure 57: Blind #2 - Highest FP*BF*

The ground truth for Blind # 2 was a HiPoint C9.

Blind #3:

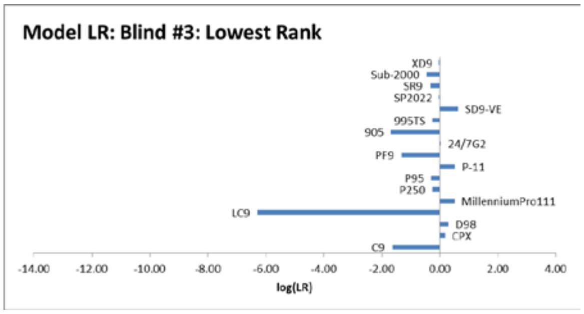

Figure 58: Blind #3 - Lowest Rank

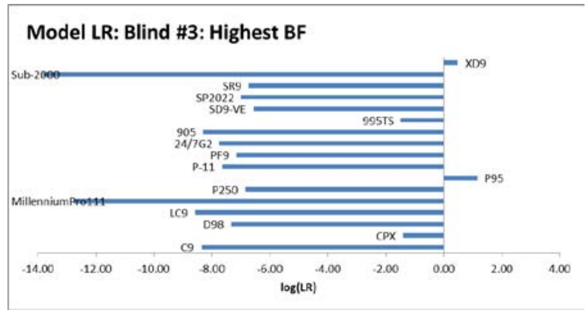

Figure 59: Blind #3 - Highest BF

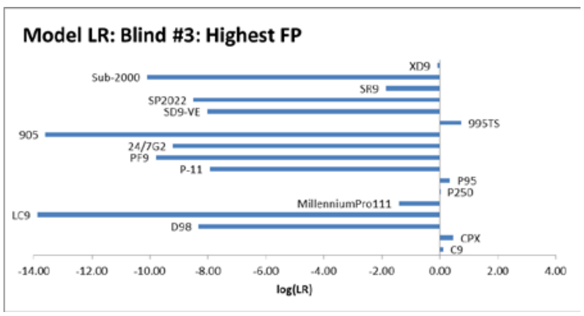

Figure 60: Blind #3 - Highest FP

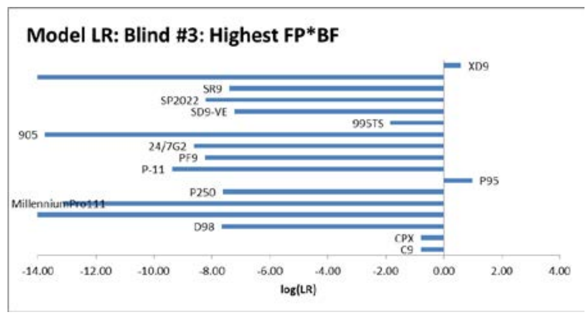

*Figure 61: Blind #3 - Highest FP*BF*

The ground truth for Blind #3 was a Springfield XD9.

Blind #4:

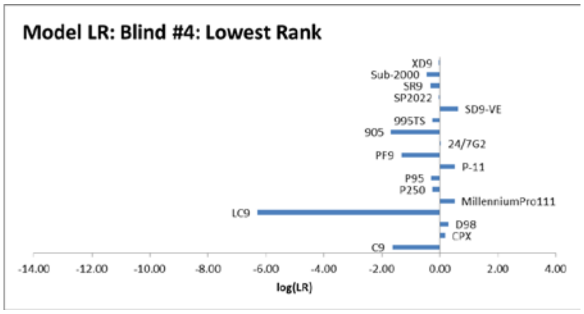

Figure 62: Blind #4 - Lowest Rank

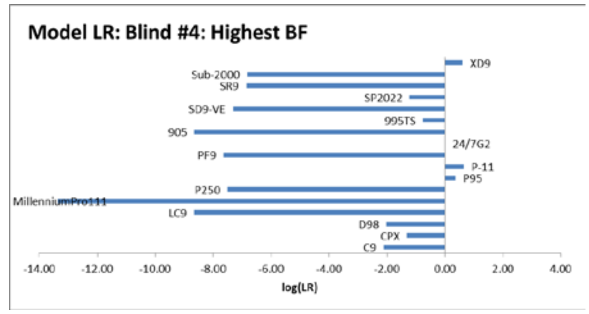

Figure 63: Blind #4 - Highest BF

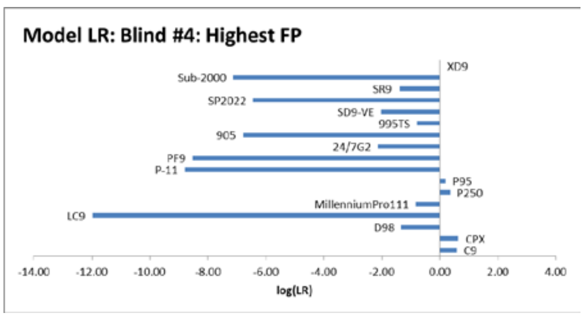

Figure 64: Blind #4 - Highest FP

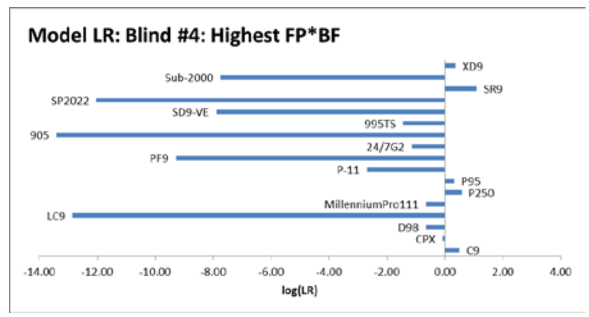

*Figure 65: Blind #4 - Highest FP*BF*

The ground truth for Blind #4 was a Glock 19. It is generally evident that the most likely make and model is a Glock from the breech face/firing pin impressions. The database has no similar cartridge cases with "Glock-type" impressions to answer this question. These results underline the general failure of this classification method.

Blind #5:

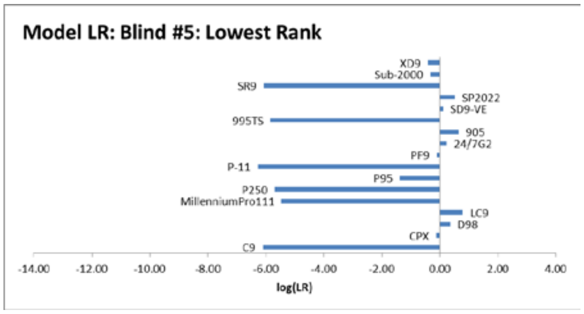

Figure 66: Blind #5 - Lowest Rank

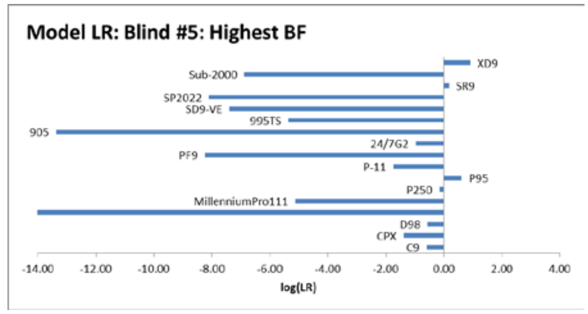

Figure 67: Blind #5 - Highest BF

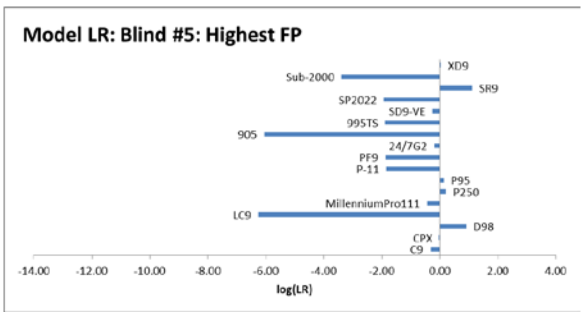

Figure 68: Blind #5 - Highest FP

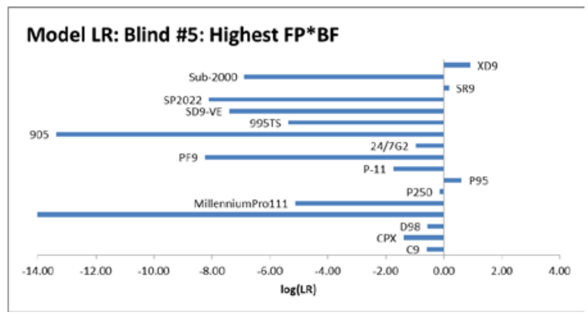

*Figure 69: Blind #5 - Highest FP*BF*

The ground truth for Blind #5 was a HiPoint 995.

Conclusion

Given these results, the ability to infer the make and/or model of a firearm from IBIS scores seems limited at present. The introduction of categorical features (such as the general class features of the breech face marks e.g. parallel, arched, cross-hatched, etc.) of the questioned cartridge case may add some additional discriminatory power. Added discrimination is needed since the scores distributions alone are too similar between the various models.

Comparison

The comparison data were evaluated in a few ways. A traditional statistical approach and a Bayesian approach were undertaken. Each set comprised of 3 knowns (K's) and one questioned (Q) cartridge cases. In Figure 70, the match (matches between K1, K2, and K3[45] (green trace)), non-match (K1, K2, K3, and Q1 versus the cartridge cases in the existing database[46] (red trace)),

[45] This describes the intra-variability of the scores of the known cartridge cases.
[46] Density of the non-match scores. Assumes that neither the known nor the questioned cartridge cases (if they are different) are represented in the database.

and unknown (Q1 versus K1, K2, and K3 (blue trace)[47]) density distributions are given for both the FP and BF scores. By inspection of the FP distributions it appears that the unknown distribution is similar to the of the non-match distribution

Set 1:

For example, in Figure 70, the match (matches between K1, K2, and K3), non-match (K1, K2, K3, and Q1 versus a pre-existing case in the database), and unknown (Q1 versus K1, K2, and K3) density distributions are given for both the FP and BF scores. By inspection of the FP distributions it appears that the unknown distribution is similar to the of the non-match distribution. This implies that the questioned cartridge case was not fired from the same firearm that fired the known cartridge cases.

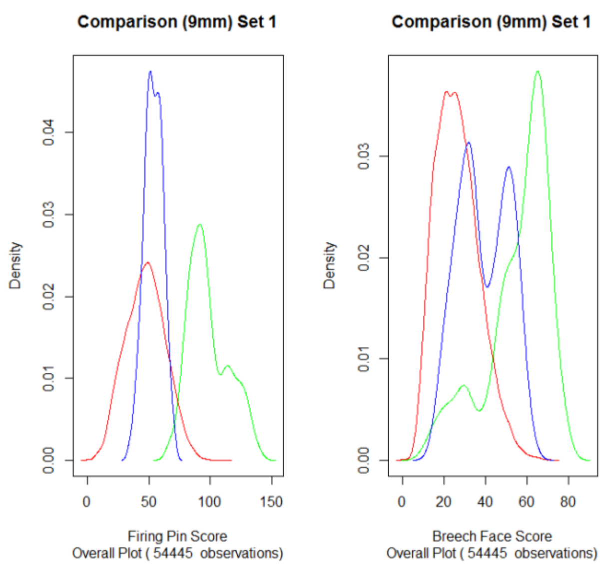

Figure 70: Comparison Set 1: FP and BF score distributions by Match type

An ANOVA was carried out on the match result and the BF and FP scores. In both cases the p value was significantly less than the significance level (0.05), thus H_0 was rejected. The pairwise comparisons and the Tukey Honest Significant Difference (HSD) test both indicated that each distribution was in its own unique set for the FP and BF mean scores (see Figure 71 and Figure 72).

---

[47] Measure of the comparison scores between Q and the set of knowns.

Set 1: Firing pin comparisons

Table 3: Set 1: Firing pin comparisons: ANOVA: FP~Match

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Match	2	2967137	1483568	5781	<2e-16 ***
Residuals	54442	13970315	257		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 4: Set 1: Firing pin comparisons: Pairwise comparisons using t tests with non-pooled SD

data: test$FP and test$Match

	No	Unknown
Unknown	<2e-16	-
Yes	<2e-16	<2e-16

P value adjustment method: holm

Table 5: Set 1: Firing pin comparisons: Tukey multiple comparisons of means: 95% family-wise confidence level: Fit: aov(formula = FP ~ Match, data = test)

$Match

	diff	lwr	Upr	p adj
Unknown-No	6.46533	4.534989	8.395671	0
Yes-No	50.92568	49.813682	52.037686	0
Yes-Unknown	44.46035	42.244621	46.676087	0

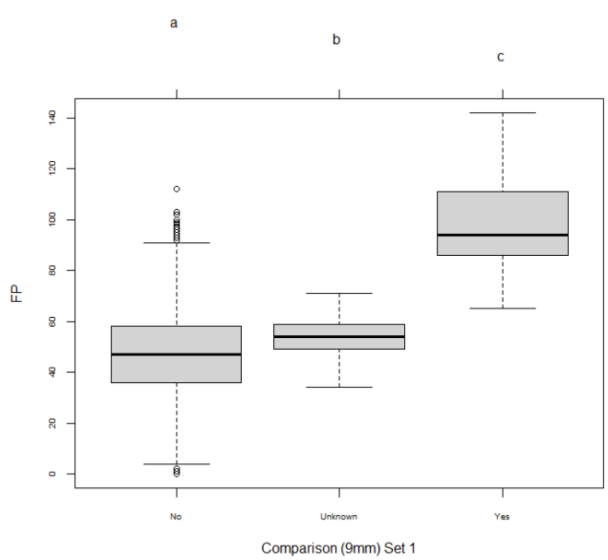

Figure 71: Comparison Set 1: Tukey HSD for firing pin scores

Set 1: Breech face comparisons

Table 6: Set 1: Breech face comparisons: ANOVA: BF~Match

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Match	2	1019315	509657	4531	<2e-16 ***
Residuals	54442	6124208	112		

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Table 7: Set 1: Breech face comparisons: Pairwise comparisons using t tests with non-pooled SD

data: test$BF and test$Match

	No	Unknown
Unknown	<2e-16	-
Yes	<2e-16	<2e-16

P value adjustment method: holm

Table 8: Set 1: Breech face comparisons: Tukey multiple comparisons of means: 95% family-wise confidence level. Fit: aov(formula = BF ~ Match, data = test)

$Match

	diff	lwr	Upr	p adj
Unknown-No	12.55417	11.27610	13.83224	0
Yes-No	29.10331	28.36706	29.83957	0
Yes-Unknown	16.54914	15.08211	18.01617	0

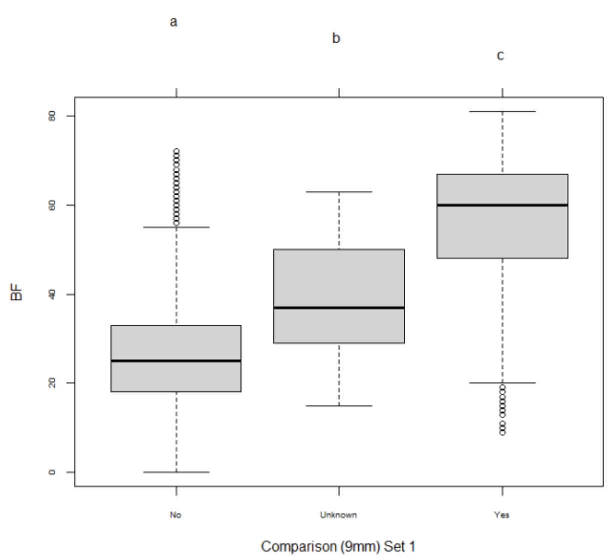

Figure 72: Comparison Set 1: Tukey HSD for breech face scores

Set 2:

For example, in Figure 73, inspection of the FP and BF distributions it appears that the unknown distribution is similar to the of the non-match distribution of FP and even more so for the BF scores. This situation can also be inferred from the plots in Figure 74 and Figure 75. There is overlap between the non-math and unknown distributions. This implies that the questioned cartridge case was not fired from the same firearm which fired the known cartridge cases.

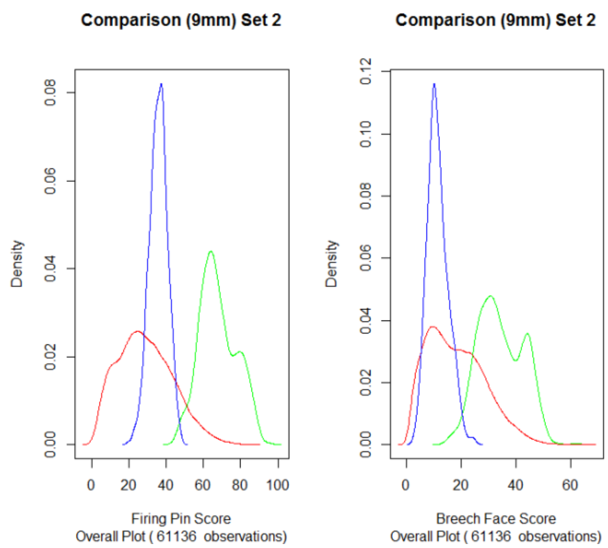

Figure 73: Comparison Set 2: FP and BF score distributions by Match type

Set 2: Firing pin comparisons

Table 9: Set 2: Firing pin comparisons: ANOVA: FP~Match

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Match	2	1752031	876015	4132	<2e-16 ***
Residuals	61133	12962120	212		

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Table 10: Set 2: Firing pin comparisons: Pairwise comparisons using t tests with non-pooled SD

data: test$FP and test$Match

	No	Unknown
Unknown	<2e-16	-
Yes	<2e-16	<2e-16

P value adjustment method: holm

Table 11: Set 2: Firing pin comparisons: Tukey multiple comparisons of means: 95% family-wise confidence level. Fit: aov(formula = FP ~ Match, data = test)

$Match

	diff	lwr	Upr	p adj
Unknown-No	6.848036	5.432505	8.263567	0
Yes-No	38.753682	37.748283	39.759082	0
Yes-Unknown	31.905646	30.180733	33.630560	0

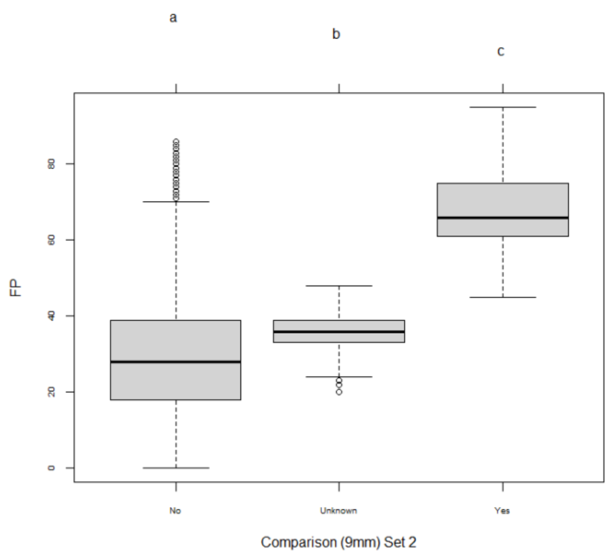

Figure 74: Comparison Set 2: Tukey HSD for firing pin scores

Set 2: Breech face comparisons

Table 12: Set 2: Breech face comparisons: ANOVA: BF~Match

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Match	2	335591	167795	1587	<2e-16 ***
Residuals	61133	6463259	106		

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Table 13: Set 2: Breech face comparisons: Pairwise comparisons using t-tests with non-pooled SD

data: test$BF and test$Match

	No	Unknown
Unknown	<2e-16	-
Yes	<2e-16	<2e-16

P value adjustment method: holm

Table 14: Set 2: Breech face comparisons: Tukey multiple comparisons of means: 95% family-wise confidence level. Fit: aov(formula = BF ~ Match, data = test)

$Match

	diff	lwr	Upr	p adj
Unknown-No	-6.670013	-7.669569	-5.670457	0
Yes-No	16.328807	15.618859	17.038755	0
Yes-Unknown	22.998820	21.780798	24.216841	0

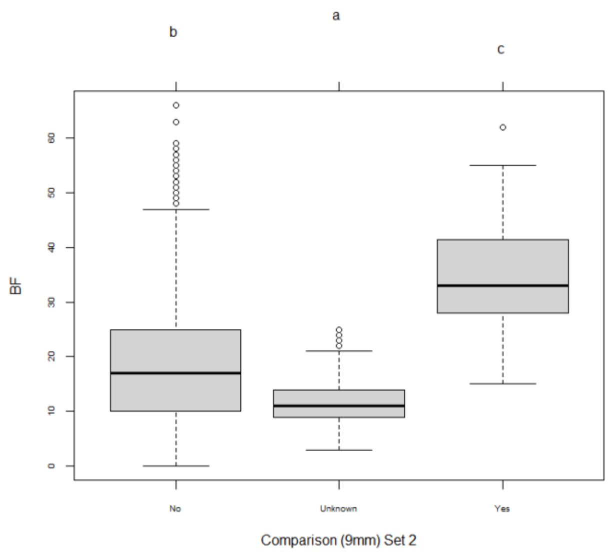

Figure 75: Comparison Set2: Tukey HSD for breech face scores

Set 3:

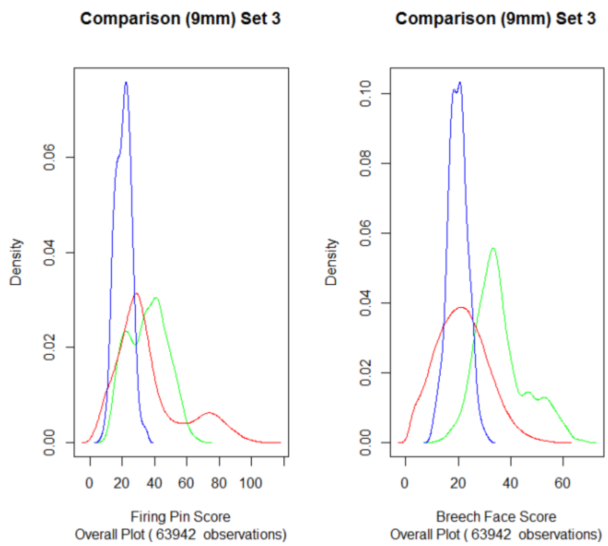

Figure 76: Comparison Set 3: FP and BF score distributions by Match type

In Figure 76, the Match, Non-Match and Unknown distributions for the FP scores overlap significantly (see also Figure 77). For the BF scores, the Match distribution is shifted slightly higher in score but the score means are still comparable to the of the BF scores. Both scores are relatively low. This implies an inconclusive result.

Set 3: Firing pin comparisons

Table 15: Set 3: Firing pin comparisons: ANOVA: FP~Match

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Match	2	84715	42357	98.04	<2e-16 ***
Residuals	63939	27623674	432		

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Table 16: Set 3: Firing pin comparisons: Pairwise comparisons using t tests with non-pooled SD:

data: test$FP and test$Match

	No	Unknown
Unknown	<2e-16	-
Yes	0.51	<2e-16

P value adjustment method: holm

Table 17: Set 3: Firing pin comparisons: Tukey multiple comparisons of means: 95% family-wise confidence level. Fit: aov(formula = FP ~ Match, data = test)

$Match

	diff	lwr	Upr	p adj
Unknown-No	-15.2766538	-17.833932	-12.719376	0
Yes-No	-0.2484222	-1.745561	1.248717	0.9200243
Yes-Unknown	15.0282316	12.077782	17.978682	0

In Set 3, the mean of firing pin scores between the match and the non-match groups cannot be differentiated.

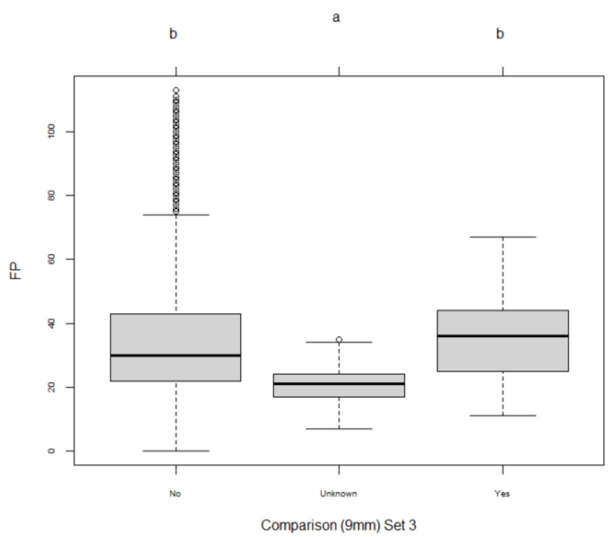

Figure 77: Comparison Set 3: Tukey HSD for firing pin scores

Set 3: Breech face comparison

Table 18: Set 3: Breech face comparisons: ANOVA: BF~Match

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Match	2	233678	116839	1225	<2e-16 ***
Residuals	63939	6099180	95		

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Table 19: Set 3: Breech face comparisons: Pairwise comparisons using t tests with non-pooled SD

data: test$BF and test$Match

	No	Unknown
Unknown	3.8e-15	-
Yes	<2e-16	<2e-16

P value adjustment method: holm

Table 20: Set 3: Breech face comparisons: Tukey multiple comparisons of means: 95% family-wise confidence level. Fit: aov(formula = BF ~ Match, data = test)

$Match

	diff	lwr	Upr	p adj
Unknown-No	-1.602882	-2.804518	-0.4012461	0.0050352
Yes-No	14.816611	14.113123	15.520100	0
Yes-Unknown	16.419494	15.033110	17.8058767	0

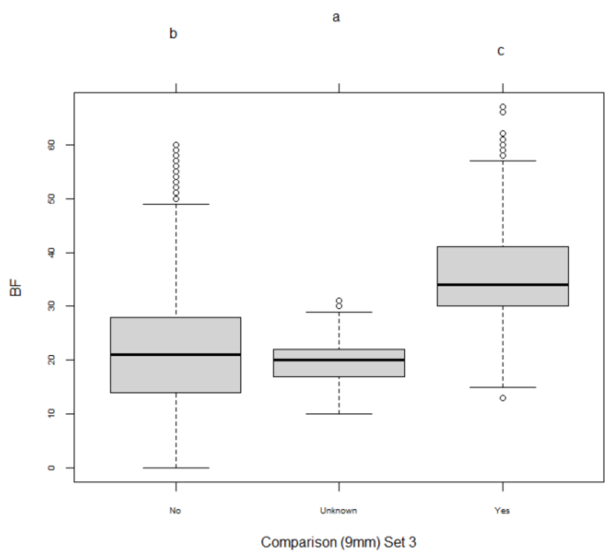

Figure 78: Comparison Set 3: Tukey HSD for breech face scores

From the Bayesian Network perspective and given the selection of the data sets to compare, the LR for each instance was computed and is given in Table 21.

Table 21: Sample LR for comparison sets

Set	Rank	BF	FP	LR	Verbal (limited evidence to support)	Ground Truth[48]	Known	Questioned
1	306	60	67	5.3	H_p	Same Gun	Ruger P95DC	Ruger P95DC
1	760	63	49	2.6	H_p	Same Gun	Ruger P95DC	Ruger P95DC
1	172	38	71	0.6	H_d	Same Gun	Ruger P95DC	Ruger P95DC
2	48	24	41	1.1	H_p	Different Gun	HiPoint C9	HiPoint 995
2	95	25	37	0.4	H_d	Different Gun	HiPoint C9	HiPoint 995
2	27	10	48	1.4	H_p	Different Gun	HiPoint C9	HiPoint 995
3	698	28	32	0.2	H_d	Different Gun	Springfield XD9	HiPoint C9
3	875	31	23	0.2	H_d	Different Gun	Springfield XD9	HiPoint C9
3	586	21	35	0.2	H_d	Different Gun	Springfield XD9	HiPoint C9

From Table 21, it can be seen that varying results were obtained with the test fire for a particular sample cartridge case. This supports the range of variabilities seen in a test set. The sample set #2 proved to be the most difficult, but overall the generated likelihood ratios proved to be very close to 1 (neutral). The absolute scores of set 2 were similar to those of set 3.

Conclusion

At the request of the program manager, the IBIS was tested under both repeatability and reproducibility conditions by a number of IBIS technicians. The standard Eurachem definitions were used to assess each condition by means of the coefficient of variation (CoV). For repeatability the maximum CoV (BF) was 9% and the maximum CoV (FP) was 28%. This variability between examiners may seem high for the FP CoV, but it must be remembered that the score values obtained in this study are extremely high scores, not usually seen in casework since the same cartridge case was used in each instance. For reproducibility the maximum CoV (BF) was 11% and the maximum CoV (FP) was 29%, being very similar to the repeatability results.

A preliminary Bayesian network was developed to assess the viability of determining the make and/or model of a firearm from the IBIS data. The results using the lowest rank, highest BF score, highest FP score, and highest BFxFP score achieved were of no significance. No further effort was expended in this direction.

A preliminary evaluation of three blind sets was carried out at the request of the program manager. In three of the nine results an incorrect attribution was made. In one case a LR of 0.6 was obtained when the ground truth was the same gun (false negative). In another set, the

[48] If the background color of the ground truth column is red, then the incorrect inference is made. If green, then the inference is correct.

ground truth of a different gun was attributed with LRs of 1.1 and 1.4 (false positive). In retrospect, these LRs are all very close to unity which implies that the evidence is neutral.

Analysis of NIST standard cartridge cases

In order to assess the standard performance of IBIS, five standard cartridge cases from NIST were used[49]. Each standard reference material (SRM) was entered into IBIS 10 times by each of three users (EBF, RLJ, and EFL). Each submission is called capture. Thus for this study each SRM was captured 30 times for a total of 150 captures. Each of these users has more than 12 months experience entering cartridge cases into IBIS.

The SRM's used were 2P2333, 2P2335, 2P2415, 2P4316, and 2P6325. These were run as normal 9 mm Luger cartridge cases and the candidate lists were processed in the usual manner.

The distribution of the data is given in Figure 79.

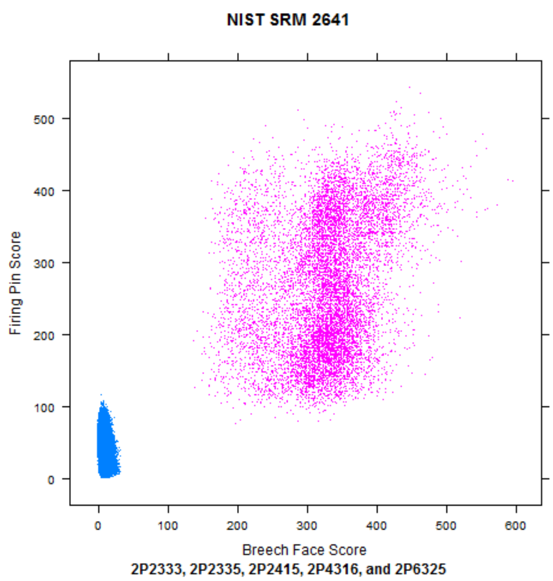

Figure 79: Firing pin vs. breech face scores for the 5 NIST SRM's (N = 299,959)

This study is one of repeated acquisitions. The extent to which each capture (SRM, repeat, and user) was found in the candidate list is given in Figure 80. Recovery values < 100% are coded in pink. For the 150 samples submitted), there are 11,175 possible comparisons (${}^{150}_{2}C$). Since all of the samples are re-correlated, there are twice the number of comparisons (a vs. b, and b vs. a) giving 22,350 comparisons. Of these, 63 comparisons were not recovered by the IBIS system. It

[49] Standard Reference Material® 2461.

is interesting that the recoveries are asymmetrical. For example, EBF-2P2415 vs. RLJ-2P2333 has a recovery of 98%, whilst RLJ-2P2333 vs. EBF-2P2415 has a recovery of 100%.

From Figure 81 it is clear that self-recovery will not occur (e.g. a. vs. a.), thus the diagonal has no instances of comparisons returned. In all cases, there will not be a recovery for the sample against itself. Therefore, on the diagonal of Figure 80, all of the values should be 90 (10x10-10). In most instances, the candidate lists yield at least 2,000 candidates. All of the lists contained non-match data. It is unclear why the recovery loss, although small (~0.28%), occurs.

		Analyst_DB																
	Match = Yes	EBF					EFL					RLJ						
		IdentifierGun_DB																
Analyst_Sample	IdentifierGun_Sample	2P2333	2P2335	2P2415	2P4316	2P6325	2P2333	2P2335	2P2415	2P4316	2P6325	2P2333	2P2335	2P2415	2P4316	2P6325	Grand Tot	% Recovery
EBF	2P2333	84	97	100	100	100	100	100	100	100	100	100	100	100	100	100	1481	99.40%
	2P2335	100	88	93	99	100	100	100	100	100	100	100	100	100	100	100	1480	99.33%
	2P2415	100	100	90	94	98	100	100	100	100	100	98	100	100	100	100	1480	99.33%
	2P4316	100	100	100	90	100	100	100	100	100	100	100	100	100	100	100	1490	100.00%
	2P6325	100	100	100	100	90	100	100	100	100	100	100	100	100	100	100	1490	100.00%
EFL	2P2333	100	100	100	100	100	90	100	100	100	100	100	100	100	100	100	1490	100.00%
	2P2335	100	100	100	100	100	100	90	98	97	98	100	100	100	100	100	1483	99.53%
	2P2415	100	100	100	100	100	100	100	90	100	100	100	100	100	100	100	1490	100.00%
	2P4316	100	100	100	100	100	100	100	100	90	100	100	100	100	100	100	1490	100.00%
	2P6325	100	100	100	100	100	100	100	100	100	90	100	100	100	100	100	1490	100.00%
RLJ	2P2333	100	100	100	100	100	100	100	100	100	100	90	100	100	100	100	1490	100.00%
	2P2335	100	100	100	100	100	98	100	100	100	100	100	87	98	99	98	1480	99.33%
	2P2415	100	100	100	100	100	100	98	99	98	96	100	100	90	100	100	1481	99.40%
	2P4316	100	100	100	100	100	100	100	100	100	100	100	100	100	82	100	1482	99.46%
	2P6325	100	100	100	100	100	100	100	100	100	100	100	100	100	100	90	1490	100.00%
	Grand Total	1484	1485	1483	1483	1488	1488	1488	1487	1485	1484	1488	1487	1488	1481	1488	22287	
	% Recovery	99.60%	99.66%	99.53%	99.53%	99.87%	99.87%	99.87%	99.80%	99.66%	99.60%	99.87%	99.80%	99.87%	99.40%	99.87%		99.72%

Figure 80: IBIS percentage recovery of NIST standards (samples = 150)

A breakdown of the missing comparisons for the subset (user=EBF, SRM=2P2333, repeat=1..10) is given in Figure 81.

Match = Yes	**Analyst_DB**	**EBF**										
	IdentifierGun_DB	**2P2333**										
		ExhibitNumber_DB										
IdentifierGun_Sample	**ExhibitNumber_Sample**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Total**
2P2333	1		1	1	1	1	1	1	1	1	1	9
	2	1			1	1	1	1	1	1	1	8
	3	1	1		1		1	1	1	1	1	8
	4	1	1	1		1		1	1	1	1	8
	5	1	1	1	1		1	1		1	1	8
	6	1	1	1	1	1		1	1		1	8
	7	1	1	1	1	1	1		1	1		8
	8	1	1	1	1	1	1	1		1	1	9
	9	1	1	1	1	1	1	1	1		1	9
	10	1	1	1	1	1	1	1	1	1		9
	Total	**9**	**9**	**8**	**9**	**8**	**8**	**9**	**8**	**8**	**8**	**84**

Figure 81: Recoveries for (user=EBF, SRM=2P2333, repeat=1..10)

Figure 82 demonstrates the performance of the IBIS system by analyst. In four of the nine comparisons perfect recovery was achieved. Of the remaining five, the average recovery was 98.30%. Interestingly, no analyst achieved a 100% recovery against their own submissions.

Match = Yes					
	Analyst_DB				
Analyst_Sample	**EBF**	**EFL**	**RLJ**	**Grand Total**	**% Recovery**
EBF	2423	2500	2498	7421	99.61%
EFL	2500	2443	2500	7443	99.91%
RLJ	2500	2489	2434	7423	99.64%
Grand Total	7423	7432	7432	22287	
% Recovery	99.64%	99.76%	99.76%		99.72%

Figure 82: IBIS percentage recovery of NIST standards (samples = 150) by Analyst

SRM	**auc.BF**	**auc.FP**	**auc.FPBF**
2P2333	1	0.999994529	1
2P2335	1	0.999999647	1
2P2415	1	0.99999956	1
2P4316	1	0.999999711	1
2P6325	1	0.999998282	1

Figure 83: Area under the Receiver operating characteristic curves for BF, FP and BFxFP scores for the SRM

The firing pin (FP) and breech face (BF) scores, as well as their product (FPBF) were evaluated according to their receiver operating characteristic (ROC) curves. The area under the curves (AUC) was measure and the results are given in Figure 83. For these data both the BF and FPBF have perfect discrimination (AUC = 1.0), whilst the FP is near perfect.

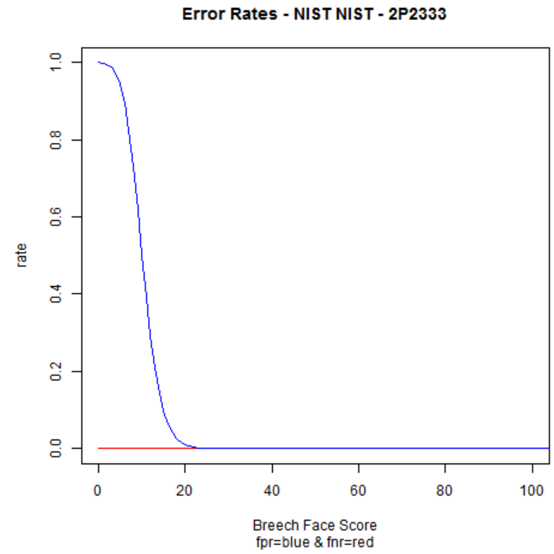

Figure 84: Error rate curve for the breech face score of 2P2333.

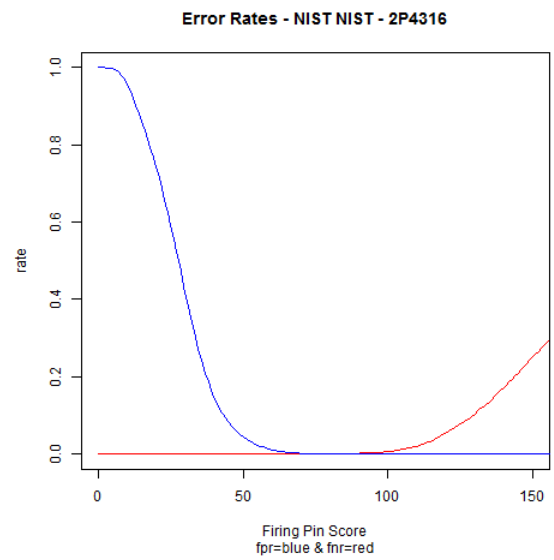

Figure 85: Error rate curve for the firing pin score of 2P4316

The utility of the BF and FP scores as a classifier for the match status of a NIST SRM is given in Figure 84 and Figure 85. In Figure 84, it can be seen that if an SRM has a BF score more than about 25 then it is a Match, without the influence of false negatives at higher scores. In Figure 85, it can be seen that if an SRM has a BF score more than about 65 then it is a match. If the score increases to about 80, then the possibility of false negatives becomes real.

Conclusion

In an attempt to assess the reliability of the IBIS results an expanded study of the NIST Standard Reference Material® 2461 was undertaken. Five of the NIST standards were tested under the same conditions as the reproducibility study. The IBIS was able to classify perfectly based on the BF score and the BFxFP score and almost perfectly on the FP score. Interestingly, the BF and FP scores between and within the standards ranged from 100 to 600.

Normalization study

During a meeting with the representatives of Ultra Electronics Forensic Technology Inc. (FTI) the concept of score normalization was discussed. The procedure for the normalization of the scores was received from FTI. Briefly, the procedure is as follows:

1. For each entered cartridge determine the number of non-matching scores ($N_{non\text{-}match}$).
2. Determine a sampling rate (SR) (10% was recommended).
3. For both firing pin and breech face scores the mean and standard deviation must be calculated using the highest $N_{non\text{-}match}$ x SR non-match (different gun) scores.
4. All scores (both match and non-match) are normalized using $z = \frac{score - mean}{std\ deviation}$
5. These normalized scores are then used as discriminate is instead of the standard score.

It was also stated that the ranks of the breech face and firing pin scores are more discriminating. Up to this point, only the rank of the firing pin has been used in calculations. According to FTI, the IBIS correlation process is broken down into two sub-processes, coarse and fine correlation. The course correlation is a fast but less accurate correlation. The objective of this process is to reject rapidly the matching candidates. This process is performed independently on the breech face and firing pin scores. The top 10% of candidates from each list all then processed using the fine correlation procedure. The scores calculated during this fine correlation process all the scores which are provided by the system. The scores calculated during the course correlation are not used further in the process. This approach can result in a candidate having a high breech face score and a low firing pin score for example. In this case, the candidate was identified through the course correlation of the breech face scores.

Table 22: R Script used for the calculation of normalized score for breech face, firing pin, and their product

FTI Score Normalization.

The file is read into the script.

```
AA9MM <- read.csv("D:/Test/AAN_00724BFR.csv")
```

As a test aid, the prefixes to the file in which data are stored are added as a categorical to the dataset. This allows for the identification of errors in a particular data file. The new file is written to the hard drive.

```
CaseID_pre <- substr(AA9MM$CaseID_Sample, 1, 3)
test <- cbind(AA9MM, CaseID_pre)
write.csv(test, "D:/Test/AAN_00724BFRclean.csv",
   row.names = FALSE)
```

The product of the breech face and firing pin scores is added as a new column into the data frame.

```
AA9MM <- test
BFFP <- AA9MM$BF * AA9MM$FP
AA9MM <- cbind(AA9MM, BFFP)
```

The main normalization process can now occur. The data set is sliced by each particular firearm and then each particular cartridge case which was test fired and entered into IBIS. A new data frame is created to contain the normalized data in addition to the existing data.

```
guns <- unique(AA9MM$IdentifierGun_Sample)
new9mm <- c()
```

A looping structure is created to extract each the dataset for each firearm and then to subset that into the match and non-match data as new data frames. An additional data frame is created to contain the results of the normalized datasets per cartridge case on a temporary basis. FTI suggested that a sampling rate of 10% be used to determine the mean and standard deviation of the non-match data (samplerate). In order to determine the number of values to be used in the calculation, the number of cartridge cases fired per gun is also required (shots). The looping structure through each of the shots is also initiated. In the loop the data for each shot is extracted from the data set of the gun (Normal_shot). This is further subset to recover the non-match data (NonMatch_shot). The number of non-match result for the particular cartridge case is determined (readings) and the sample size to use for the calculation is determined (size). In order to calculate the mean and standard deviations, at least three values are required in the size sample. These tests check to determine whether these requirements are met. The calculation for the normalization of the breech face scores was undertaken. The dataset is sorted in decreasing BF score. The top samplerate% of the readings are then assigned to the TS vector. The mean and standard deviation of this vector are determined to calculate the new BF-norm vector for the cartridge case. A similar approach is used for both the FP scores and the BFFP scores. The BF-norm, FP_norm, BFFP_norm values are added to the data frame for the cartridge case. All of the cartridge cases for the particular gun are then assigned to the data frame New. The combination of gun and cartridge case is printed to assess progress in the conversion. The data for each completed gun is then attached to the overall data frame new9mm.

```
for (j in guns) {
  Normal <- subset(AA9MM, IdentifierGun_Sample == j, select = c(CaseID_Sample,
    ExhibitNumber_Sample, Rank, CaseID_DB, ExhibitNumber_DB, BF, FP, Match,
    Make_DB, Model_DB, Ammo_DB, Caliber_DB, Firing_Pin_Type_DB, Make_Sample,
    Model_Sample, Ammo_Sample, IdentifierGun_Sample, IdentifierGun_DB, Caliber_Sample,
    Firing_Pin_Type_Sample, Type_Sample, Primer_Sample, Primer_DB, Drag_Mark_Sample,
    Drag_Mark_DB, Reload, Rank_BF, BFFP, CaseID_pre))

  NonMatch <- subset(Normal, Match == "No", select = c(CaseID_Sample, ExhibitNumber_Sample,
    Rank, CaseID_DB, ExhibitNumber_DB, BF, FP, Match, Make_DB, Model_DB,
    Ammo_DB, Caliber_DB, Firing_Pin_Type_DB, Make_Sample, Model_Sample,
    Ammo_Sample, IdentifierGun_Sample, IdentifierGun_DB, Caliber_Sample,
    Firing_Pin_Type_Sample, Type_Sample, Primer_Sample, Primer_DB, Drag_Mark_Sample,
    Drag_Mark_DB, Reload, Rank_BF, BFFP, CaseID_pre))

  Match <- subset(Normal, Match == "Yes", select = c(CaseID_Sample, ExhibitNumber_Sample,
    Rank, CaseID_DB, ExhibitNumber_DB, BF, FP, Match, Make_DB, Model_DB,
    Ammo_DB, Caliber_DB, Firing_Pin_Type_DB, Make_Sample, Model_Sample,
    Ammo_Sample, IdentifierGun_Sample, IdentifierGun_DB, Caliber_Sample,
    Firing_Pin_Type_Sample, Type_Sample, Primer_Sample, Primer_DB, Drag_Mark_Sample,
    Drag_Mark_DB, Reload, Rank_BF, BFFP, CaseID_pre))

  New <- c()

  samplerate = 0.1

  shots <- unique(Match$ExhibitNumber_Sample)

  for (i in shots) {
    Normal_shot <- subset(Normal, ExhibitNumber_Sample == i, select = c(CaseID_Sample,
      ExhibitNumber_Sample, Rank, CaseID_DB, ExhibitNumber_DB, BF, FP,
      Match, Make_DB, Model_DB, Ammo_DB, Caliber_DB, Firing_Pin_Type_DB,
      Make_Sample, Model_Sample, Ammo_Sample, IdentifierGun_Sample, IdentifierGun_DB,
      Caliber_Sample, Firing_Pin_Type_Sample, Type_Sample, Primer_Sample,
      Primer_DB, Drag_Mark_Sample, Drag_Mark_DB, Reload, Rank_BF, BFFP,
      CaseID_pre))
    NonMatch_shot <- subset(NonMatch, ExhibitNumber_Sample == i, select = c(CaseID_Sample,
      ExhibitNumber_Sample, Rank, CaseID_DB, ExhibitNumber_DB, BF, FP,
      Match, Make_DB, Model_DB, Ammo_DB, Caliber_DB, Firing_Pin_Type_DB,
      Make_Sample, Model_Sample, Ammo_Sample, IdentifierGun_Sample, IdentifierGun_DB,
      Caliber_Sample, Firing_Pin_Type_Sample, Type_Sample, Primer_Sample,
      Primer_DB, Drag_Mark_Sample, Drag_Mark_DB, Reload, Rank_BF, BFFP,
      CaseID_pre))

    readings <- length(NonMatch_shot$BF)
    size <- ceiling(readings * samplerate)

    if (size < 0.1) {
      cat("zero non matches check for error")
    }
    if (size < 0.2) {
      cat("one non match delete cartridge")
    }
```

```
    if (size <= 0.3) {
      size <- readings
    }

    # BF Norm
    sortBF <- sort(NonMatch_shot$BF, decreasing = T)
    TS <- head(sortBF, size)
    TS_ave <- mean(TS)
    TS_sd <- sd(TS)
    BF_norm <- (Normal_shot$BF - TS_ave)/TS_sd

    # FP Norm
    sortFP <- sort(NonMatch_shot$FP, decreasing = T)
    TS <- head(sortFP, size)
    TS_ave <- mean(TS)
    TS_sd <- sd(TS)
    FP_norm <- (Normal_shot$FP - TS_ave)/TS_sd

    # BFFP Norm
    sortBFFP <- sort(NonMatch_shot$BFFP, decreasing = T)
    TS <- head(sortBFFP, size)
    TS_ave <- mean(TS)
    TS_sd <- sd(TS)
    BFFP_norm <- (Normal_shot$BFFP - TS_ave)/TS_sd

    Normal_shot <- cbind(Normal_shot, BF_norm, FP_norm, BFFP_norm)
    New <- rbind(New, Normal_shot)
    cat(j, i)
  }

  new9mm <- rbind(new9mm, New)

}

## XXX724 901XXX724 902XXX724 … XXX724 1000
```

This result is written to a file for further analysis.

```
write.csv(new9mm, "D:/Test/AAN_00724BFRclean_Normalized.csv",
  row.names = FALSE)
```

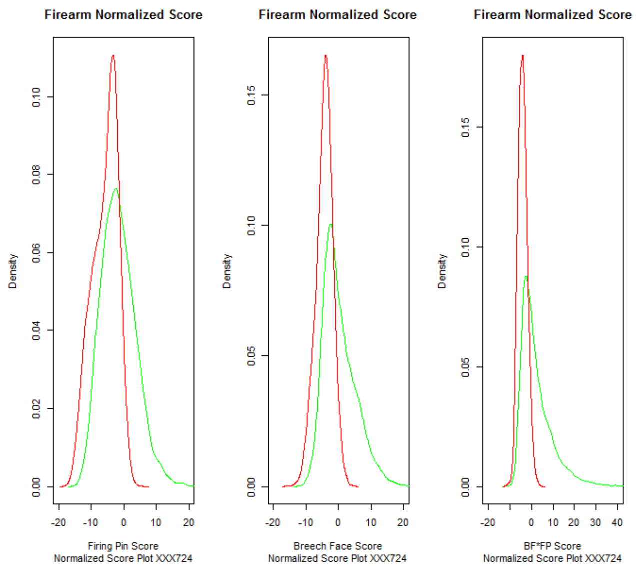

Figure 86: Density distributions of normalized firing pin, breech face, and BFFP scores from file AAN_00724 (N= 121618 observations)

When "printing" the files from IBIS, the ranks for either breech face score or firing pin score can be selected, but not both. Thus far, the firing pin rank has been used as the standard. Since rank is related to each cartridge and the data files generated contain all of the cartridge cases for a particular string, the file must be parsed to determine the breech face rank within each file.

Table 23: R Script used for the calculation of breech face rank for data files.

Calculation of breech face rank for data files.
The file is read into the script.

```
NineMM <- read.csv("D:/Test/AAN_00724.csv")
```

The primary sort of each data file is by the CaseID_Sample (each cartridge case entered in to IBIS) in increasing order. The secondary sort is by BF score in descending order.

```
newdata <- NineMM[order(NineMM$CaseID_Sample, -NineMM$BF), ]
```

The unique values of the CaseID Sample are extracted. A new dataframe, NineMMup, is created which will contain all of the existing data and the breech face rank.

```
zNewID <- unique(newdata$CaseID_Sample)
NineMMup <- c()
len <- length(zNewID)
i <- 1
```

A loop structure is used to move through the file by each unique CaseID Sample. The main file is then subset on each of the samples and the Rank_DB values are added.

```
for (i in zNewID) {
  Rank_BF <- c()
  Sample <- i

  sort_sample <- subset(newdata, CaseID_Sample == Sample, select = c(CaseID_Sample,
    ExhibitNumber_Sample, Rank, CaseID_DB, ExhibitNumber_DB, BF, FP, Match,
    Make_DB, Model_DB, Ammo_DB, Caliber_DB, Firing_Pin_Type_DB, Make_Sample,
    Model_Sample, Ammo_Sample, IdentifierGun_Sample, IdentifierGun_DB, Caliber_Sample,
    Firing_Pin_Type_Sample, Type_Sample, Primer_Sample, Primer_DB, Drag_Mark_Sample,
    Drag_Mark_DB, Reload))

  sort_sample_length <- length(sort_sample$BF)
  Rank_BF <- seq(1, sort_sample_length, by = 1)
  sort_sample <- cbind(sort_sample, Rank_BF)

  NineMMup <- rbind(NineMMup, sort_sample)
  cat(i, "\n")
}
## AAN-UK-SSG-0313-0901
## AAN-UK-SSG-0313-0902
## AAN-UK-SSG-0313-0903
.
.
.
## AAN-UK-SSG-0313-1000
```

The new data frame is written to a .csv file.

```
write.csv(NineMMup, "D:/Test/AAN_00724BFR.csv",
  row.names = FALSE)
```

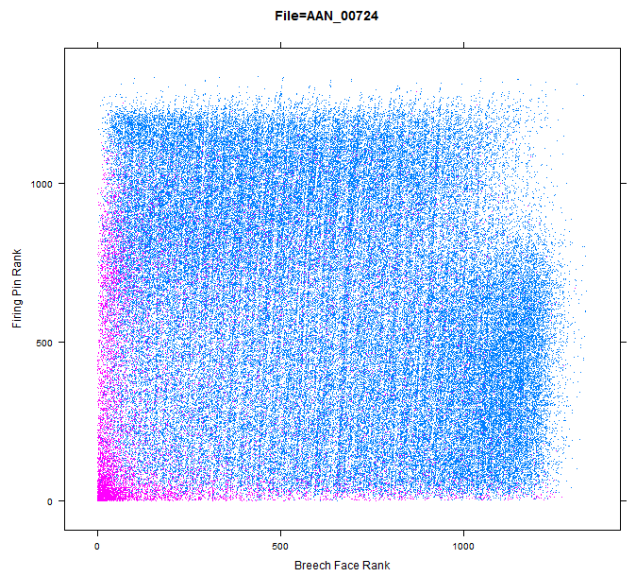

Figure 87: Firing pin rank vs. breech face rank from file AAN_00724 (N= 121618 observations).

Figure 87 presents the breech face and firing pin ranks of the Arcus D98 pistol (121,618 observations). When considering the results from Table 2 (AUC_BF = 0.718, AUC_FP = 0.717, and AUC_FPxBF = 0.786) it is seen that both the firing pin score and the breech face score perform equally well as classifiers. Figure 87 has a concentration of matching scores (pink dots) with high ranks (low values) for both firing pin and breech face. There are also bands across the axes at high ranks for each, but low ranks for the other. In the bulk of the data there are both match and non-match data at relatively low ranks.

The effect of normalization is demonstrated using a Remington R1 (1911) in .45 ACP. The normalization of the scores was undertaken using the sample rate of 10% as specified by Ultra Electronics Forensic Technology Inc. (FTI). The main purpose was to evaluate if improved discrimination between same source and different source cartridge cases was obtained. The measure of discrimination used is the area under the receiver operating characteristic curve (ROC, AUC).

Table 24: Raw and derived metrics for breech face and firing pin scores used to evaluate method efficacy

Measure	**Calculation**
auc.BF	$Raw\ BF\ score$
auc.FP	$Raw\ FP\ score$
auc.BFFP	$BF \times FP\ score$
auc.FP_Rank	$FP\ Rank$
auc.BF_Rank	$BF\ Rank$
auc.BFFP_Rank	$FP\ Rank\ \times Rank\ BF$
auc.BFFP_Score_Over_BFFP_Rank	$\frac{BF \times FP\ score}{BF\ Rank \times FP\ Rank}$
auc.BF_norm	$z_{BF} = \frac{BF\ Score - BF_{non-match}mean}{BF_{non-match}std\ deviation}$
auc.FP_norm	$z_{FP} = \frac{FP\ Score - FP_{non-match}mean}{FP_{non-match}std\ deviation}$
auc.BFFP_norm	$z_{BF} = \frac{BF \times FP\ Score - BF \times FP_{non-match}mean}{BF \times FP_{non-match}std\ deviation}$

The metrics referred to Table 24 were used to assess their applicability as a classifier of the IBIS data. In general, the scores can be categorized into three categories namely, firing pin related scores, breech face related scores, and combinations of firing pin and breech face scores. The rank scores are associated with the particular metric since ranks are simply the rank order of those scores.

Also during the study, the breech face ranks of each cartridge were determined. Generally, the IBIS system only allows one rank to be extracted. Given the data processing, it is possible to compare both raw scores and ranks to each other. For each firearm, a number of evaluations of the data were performed. Receiver operating characteristic curves and their associated error rate curves were generated for each category (breech face, firing pin, combinations) by their raw score, normalized score, and ranks.

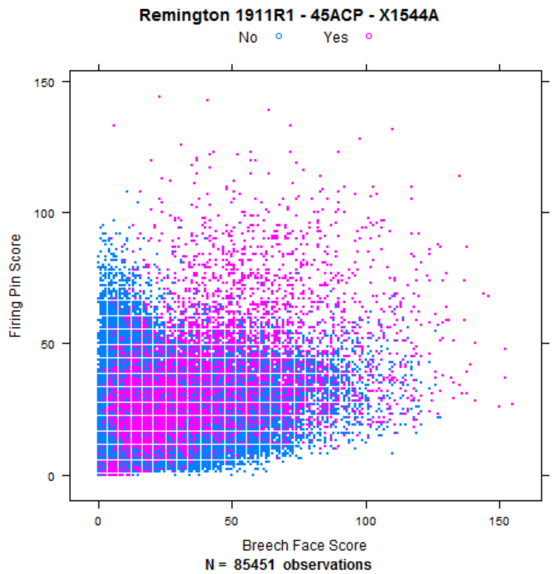

Figure 88: Scatterplot of firing pin scores versus breech face scores for a Remington model 1911 R1 in.45ACP. Scores are grouped by match status (yes/no). Sample size for this plot is 84,451 observations

The distribution of breech face and firing pin scores for a Remington model 1911 R1 in.45ACP (IdentifierGun = X1544A) is given in Figure 88. This plot indicates the separation and overlap between same source and different source guns. All different source guns are of the same caliber and of multiple makes and models (to include examples of the same make and model). When one considers the non-match distribution it can be seen that the firing pin scores reach a maximum value of just under 100, whilst the breech face scores reach a maximum at approximately 125. There is significant overlap of match and non-match scores in this region with a strong cluster at very low scores, and a high-density cluster centered around (25, 25). As one moves up the diagonal when reaching the (50, 50) position, the matches seem to separate from the non-matches. However, the density of matches in this region appears significantly lower.

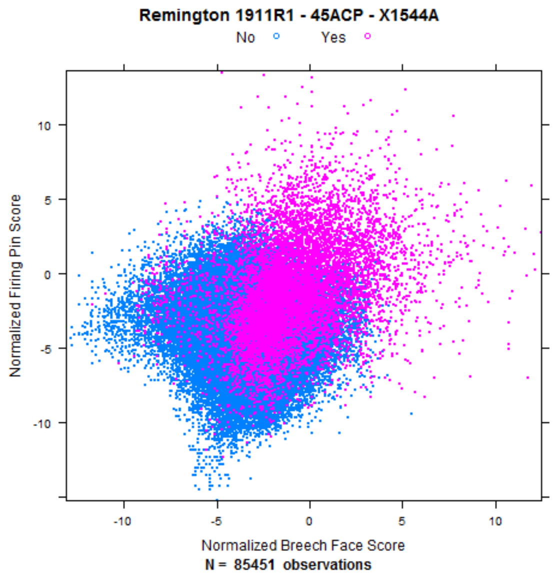

Figure 89: Scatterplot of normalized firing pin scores versus normalized breech face scores for a Remington model 1911 R1 in.45ACP. Scores are grouped by match status (yes/no). Sample size for this plot is 84,451 observations. Sample rate for normalization is 10%.

Figure 89 is a plot of the same data as in Figure 88 except that the scores of the breech face and firing pin have been normalized (using a sample rate of 10%). The normalized scores are given in terms of *z*-values (number of standard deviations from the mean). There is some interesting structure in this plot. The data seem to tail towards *z*-values of -5 for both the non-match breech face and firing pin normalized scores. There is some propensity for the same behavior for the match scores, but more so for the normalized firing pin scores. It appears that there is an improved separation between the match and non-match score for the normalized values than in the case of the raw scores, for this particular firearm.

The normalization process can be intuitively understood as a transformation process, but a qualitative description may be useful. It is important to remember that the normalization process focuses on a single cartridge case at a time. This cartridge case is searched against the database yield its respective candidate list. The candidate list, in most cases, will contain both matching and non-matching cartridge cases. It is also been observed that the scores between candidate lists sometimes demonstrate a difference. The normalization process helps in removing these differences. It provides a new score, which allows for better comparison of scores between the different cartridge cases from a particular firearm. The second feature of the normalization process allows for improved discrimination of scores between matching and non-matching cartridge cases. In the ideal case, the distributions of match and non-match scores will be completely separated (AUC = 1.0). Improvement in these two facets may be measured by evaluating the improvement in the AUC versus that of the raw scores. Forensic Technologies

Inc. indicated that they had tested sample rates of 10% and 100%. They decided to make use of the 10% sample. All of these studies were therefore performed at the same sample rate.

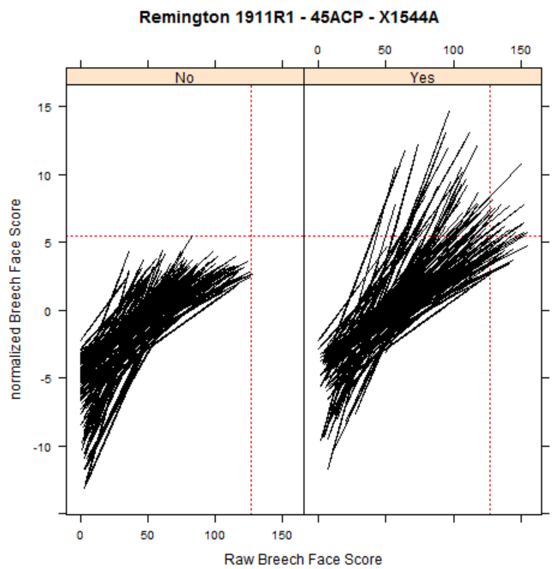

Figure 90: Plot of normalized breech face scores versus raw breech face scores for a Remington model 1911 R1 in.45ACP. Scores are grouped by match status (yes/no). Sample size for this plot is 84,451 observations. Sample rate for normalization is 10%. The red dotted lines indicate the maximum raw and normalized scores. This indicates the improvement by comparing the density to the right of the vertical red line (BF=108) to that above the horizontal red line (zBF=5.2).

In Figure 90 the breech face scores indicate that even after normalization there is still significant overlap between the non-match and match scores (confer Figure 89).

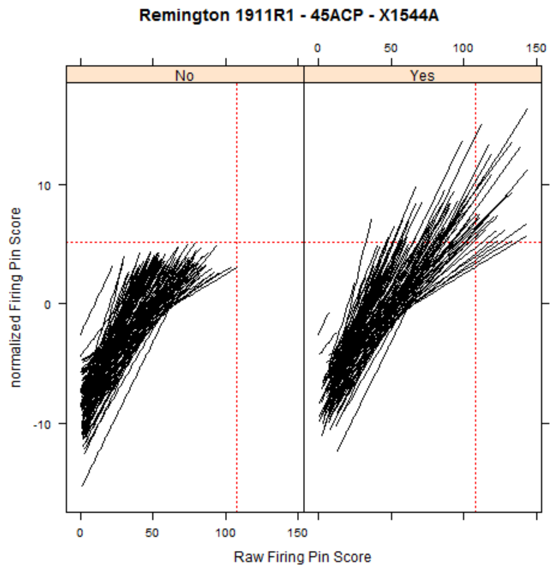

Figure 91: Plot of normalized firing pin scores versus raw firing pin scores for a Remington model 1911 R1 in.45ACP. Scores are grouped by match status (yes/no). Sample size for this plot is 84,451 observations. Sample rate for normalization is 10%. The red dotted lines indicate the maximum raw and normalized scores. This indicates the improvement by comparing the density to the right of the vertical red line (FP=127) to that above the horizontal red line (zFP=5.5).

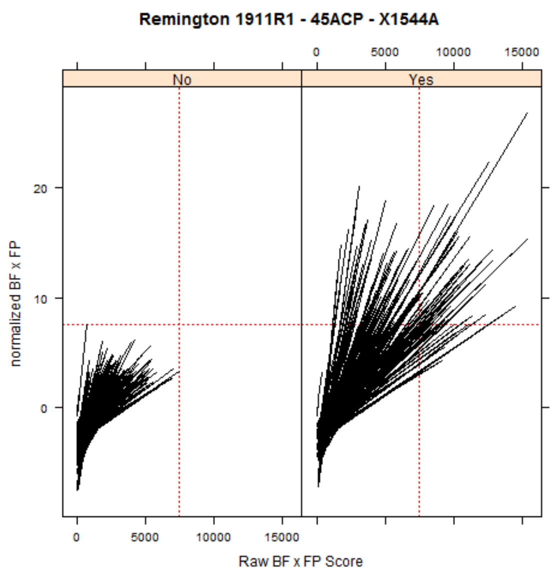

Figure 92: Plot of the product of the normalized firing pin and breech face scores versus the product of the raw firing pin and breech face scores for a Remington model 1911 R1 in.45ACP. Scores are grouped by match status (yes/no). Sample size for this plot is 84,451 observations. Sample rate for normalization is 10%. The red dotted lines indicate the maximum raw and normalized scores. This indicates the improvement by comparing the density to the right of the vertical red line (BFFP=7500) to that above the horizontal red line (zBFFP=7.6).

In Figure 90 to Figure 92, each line represents the normalized score (y-axis) obtained from the raw score (x-axis) for a given sample cartridge case. In this dataset (1911 R1 in .45ACP – X1544A) there were 119 sample cartridge cases submitted to IBIS. Thus in each plot there will be 119 transformation lines. The variation in the gradient and intercepts of the of the transformation lines indicates the variability in the collected data according to this normalization technique.

Table 25: AUC for a Remington model 1911 R1 in.45ACP (X1544A). The measures are those specified in Table 24. The AUC – Normalized values are calculated after normalization with a sample rate of 10%. The change in AUC is also given (improvement are indicated by a positive value).

X1544A			
Measure	**AUC - Raw**	**AUC - Normalized**	**% Change**
auc.BF	0.784	0.798	1.76%
auc.FP	0.739	0.754	2.07%
auc.BFFP	0.830	0.848	2.16%
auc.FP_Rank	0.761		
auc.BF_Rank	0.818		
auc.BFFP_Rank	0.844		
auc.BFFP_Score_Over_BFFP_Rank	0.845		

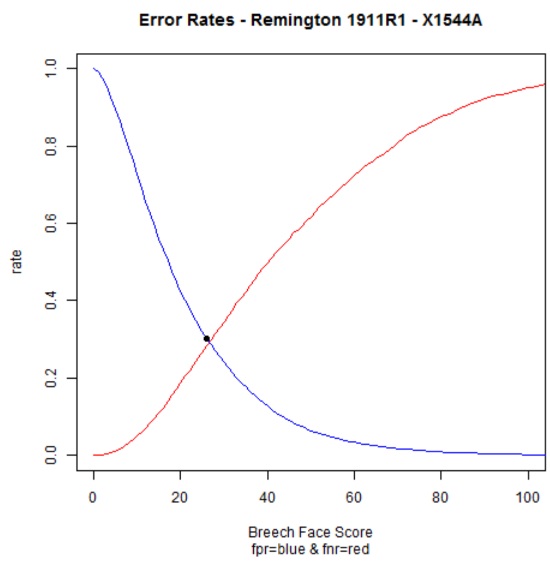

Figure 93: Error rates (FPR and FNR) versus cut-offs for the breech face scores for a Remington model 1911 R1 in.45ACP. Sample size for this plot is 84,451 observations. From Figure 48 the indicated cut-off for a zero FPR is BF=108. At this score, the FNR is significant. The black dot approximately the crossover is the EER (equal error rate).

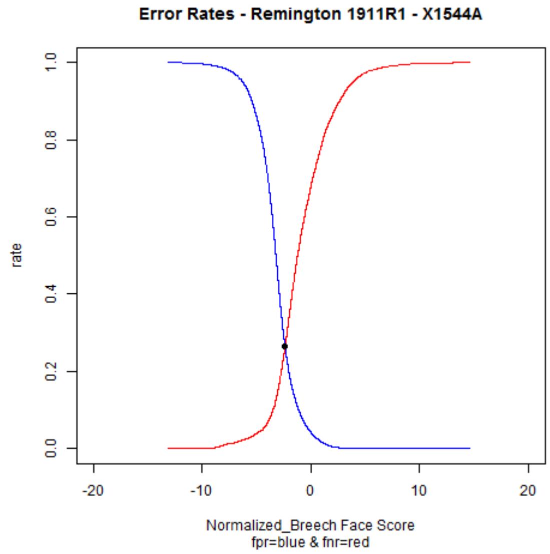

Figure 94: Error rates (FPR and FNR) versus cut-offs for the normalized breech face scores for a Remington model 1911 R1 in.45ACP. Sample size for this plot is 84,451 observations. From Figure 48 the indicated cut-off for a zero FPR is zBF=5.2. At this score, the FNR is significant although slight lower than that of the raw BF score. The black dot approximately the crossover is the EER (equal error rate).

In Figure 93 and Figure 94, the error rate curves are given that were derived from the respective ROC curves. The equal error rate (EER) is displayed as a black dot. The EER is that point where the false positive rate (FPR) equals the false negative rate (FNR). This occurs at a certain cutoff with a certain rate. The EERs of the various measures are given in Table 26.

In these plots the EER was estimated as follows:

Equation 5: Determination of the equal error rate (EER)

Choose two successive points (x and x+1) for the cut-off. If the following conditions are true, then the EER may be estimated:

If,

$$FPR(x) > FNR(x) \text{ and } FPR(x+1) < FNR(x+1)$$

then, if $EER = (i, j)$:

$$i = \frac{x + (x+1)}{2}$$

and,

$$j = \frac{FPR(x) + FNR(x) + FPR(x+1) + FNR(x+1)}{4}$$

Table 26: Equal error rates (EER) for the various measures (see Table 24). Remington model 1911 R1 in.45ACP. Sample size for this plot is 84,451 observations. The rates (fpr=fnr) are sorted in ascending (worsening) order. The absolute values of the cut-offs cannot be compared directly with each other.

X1544A	EER	
Measure	**Cut-off**	**Rate**
BFFP_norm	-1.610	22.6%
BFFP_Score_Over_BFFP_Rank	1.017E-02	22.8%
BFFP_Rank	1.840E-05	23.6%
BF_Rank	4.673E-03	24.1%
BFFP	594	24.3%
BF_norm	-2.382	26.4%
BF	26	30.2%
FP_Rank	3.937E-03	31.3%
FP_norm	-2.869	31.8%
FP	27	36.1%

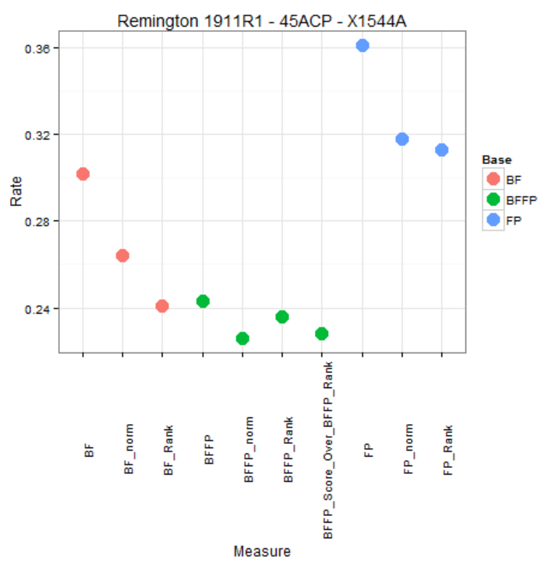

Figure 95: Equal error rates for each measure (Table 5) grouped by the three bases (BF, FP, and BFFP) for a Remington model 1911 R1 in.45ACP. The lower the rate the better the performance of the measure. Sample size for this plot is 84,451 observations.

From Figure 95, it can be seen that the different bases of the measures (BF, FP, and BFP) do not perform equally well. In general, it can be observed that discrimination of the order:

BFFP > BF > FP

for this particular firearm. For the BF and FP the order of performance is:

Rank > normalized score > raw score.

For the combined score category the order of performance is:

normalized score > combination of score and rank > rank > raw score.

The normalized scores were then used to determine their effectiveness as a classifier for match/non-match. A receiver operating characteristic (ROC) curve was then constructed to illustrate the effectiveness of the classification. The area under the ROC curve (AUC) was calculated to be used as a measure of classification effectiveness. The AUC for the raw data was used as the reference for evaluating the normalized scores. Normalization took place for the following sample rates: 0.01, 0.02, 0.03, 0.05, 0.08, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.5, 0.65, and 0.8. At each of these rates, the AUC was calculated for each of the classifiers. The normalized data are given in Table 27.

Table 27: Portion of .45 ACP data of AUC for pistols at various sample rates. The AUCs are given for both the raw and normalized scores (BF, FP, and BFFP)

Sample Rate	Make Sample	Model Sample	Identifier Gun Sample	AUC BF	AUC FP	AUC BFFP	AUC BF (norm)	AUC FP (norm)	AUC BFFP (norm)
0.01	HiPoint	JHP	X11599	0.980	0.923	0.983	0.975	0.897	0.972
0.01	Ruger	SR45	X12242	0.610	0.865	0.787	0.652	0.844	0.766
0.01	Ruger	SR45	X12243	0.574	0.810	0.745	0.602	0.792	0.715
0.01	Ruger	SR45	X12246	0.558	0.830	0.734	0.591	0.804	0.729
0.01	Remington	1911 R1	X1544A	0.784	0.739	0.830	0.709	0.719	0.773
0.01	Remington	1911 R1	X1553A	0.840	0.870	0.893	0.779	0.872	0.848
0.01	Rock Island Armory	1911 A1	X18282	0.867	0.824	0.924	0.855	0.779	0.893
0.01	Taurus	PT 145 Pro	X25098	0.744	0.825	0.857	0.744	0.817	0.837
0.01	Remington	1911 R1	X3208A	0.669	0.880	0.856	0.670	0.849	0.828
0.01	HiPoint	JHP	X45161	0.918	0.742	0.911	0.922	0.718	0.899
0.01	HiPoint	JHP	X45162	0.839	0.704	0.864	0.834	0.686	0.848
0.01	HiPoint	JHP	X45163	0.881	0.869	0.939	0.867	0.840	0.907
0.01	HiPoint	JHP	X45164	0.810	0.759	0.861	0.815	0.737	0.831
0.01	Taurus	24/7 G2	X54290	0.721	0.921	0.909	0.700	0.905	0.877
0.01	Taurus	24/7 G2	X54307	0.872	0.660	0.844	0.876	0.653	0.843
0.01	Taurus	24/7 G2	X54308	0.957	0.814	0.961	0.964	0.784	0.964
0.01	Taurus	24/7 G2	X54309	0.671	0.688	0.778	0.642	0.661	0.708
0.01	Taurus	24/7 G2	X58243	0.773	0.852	0.860	0.766	0.797	0.825
0.01	Kahr	CW 45	XE7570	0.949	0.914	0.971	0.928	0.906	0.955
0.01	Remington	1911 R1	XX544A	0.956	0.914	0.960	0.940	0.909	0.948
0.01	Taurus	24/7 G2	XX7938	0.726	0.573	0.731	0.694	0.592	0.693
0.01	Taurus	24/7 G2	XX7941	0.621	0.636	0.688	0.597	0.602	0.630
0.01	Taurus	24/7 G2	XX7943	0.608	0.775	0.752	0.592	0.791	0.722
0.01	Taurus	24/7 G2	XX7944	0.757	0.574	0.735	0.732	0.609	0.712

This data file is processed to calculate the percentage change in AUC for the various sample rates and the data are plotted as a function of the firearm make, model, and identifier (reduced serial number). A portion of the script developed to perform these calculations is given in Table 28.

Table 28: Script used to calculate the percentage change in AUC and to plot these changes

```
library(lattice)

#Load and Cleanup File=========================================
#45 ACP
samplings <- read.csv("Z:/Output Samplings/45ACP Output_Samplings/AA45ACP_samplings.csv")
title<-"Normalization by Sample Rate - 45 ACP"
vert<-.40
total<-length(samplings$auc.BF)

per_auc_BF_norm<-c()
per_auc_FP_norm<-c()
per_auc_FPBF_norm<-c()

for (i in 1:total){
  per_auc_BF_norm[i]<-((samplings$auc.BF_norm[i]-samplings$auc.BF[i])/samplings$auc.BF[i])*100
  per_auc_FP_norm[i]<-((samplings$auc.FP_norm[i]-samplings$auc.FP[i])/samplings$auc.FP[i])*100
  per_auc_FPBF_norm[i]<-((samplings$auc.FPBF_norm[i]-
      samplings$auc.FPBF[i])/samplings$auc.FPBF[i])*100
  cat(i,per_auc_BF_norm[i],per_auc_FP_norm[i],per_auc_FPBF_norm[i],"\n")
}

samplings<-cbind(samplings,per_auc_BF_norm,per_auc_FP_norm,per_auc_FPBF_norm)
horiz<-0

# IdentifierGun_Sample===================
xyplot(samplings$auc.BF_norm ~ samplings$SampleRate|samplings$IdentifierGun_Sample, par.strip.text = 
list(cex = 0.75), ylab="AUC of normalized BF ROC curve", xlab="Normalization sample rate",type="l", 
main=title, sub="by IdentifierGun_Sample")

xyplot(samplings$per_auc_BF_norm~samplings$SampleRate|samplings$IdentifierGun_Sample, par.strip.text = 
list(cex = 0.75), ylab="% change in AUC of normalized BF ROC curve", xlab="Normalization sample 
rate",type="l", main=title, sub="by IdentifierGun_Sample", panel = function(...) {
  panel.abline(h=horiz, v=vert, lty = "dotted", col = "black")
  panel.xyplot(...)
})
```

When evaluating the AUC values, it is important to note that a value of 1.0 implies a perfect classification, whilst a value of 0.5 implies that the method is equivalent to a coin toss. If an AUC of less than 0.5 is obtained, then the classification scheme should be reversed.

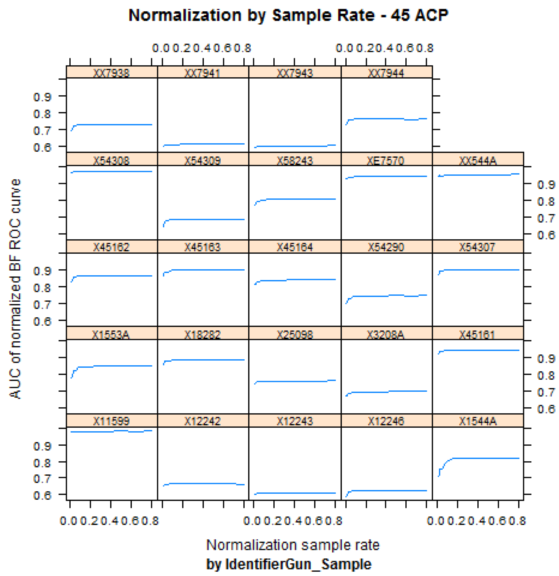

Figure 96: Change in AUC as a function of normalization sample rate for 45 ACP handguns for normalized breech face scores. A separate plot for each individual handgun is given

From Figure 96, it can be seen most of the firearms indicate a steady increase in AUC up to a sample rate of about 20%. Thereafter, the AUC stabilizes as the sample rate increases.

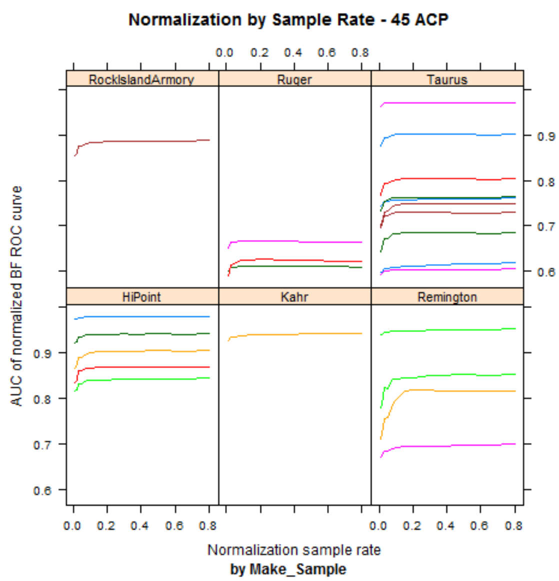

Figure 97: Change in AUC as a function of normalization sample rate for 45 ACP handguns for normalized breech face scores. A separate plot for each make of handgun is given.

In Figure 97, the same data are plotted by the make of the handgun. This demonstrates the variability in AUC with in the guns are of the same make. The change in AUC appears to be relatively similar within each make, whilst the raw AUC values are quite different. For the Taurus firearms, the AUC's vary from about 0.6 to over 0.95. The scores for the Ruger firearms are relatively low, whilst those for the HiPoint's are relatively high.

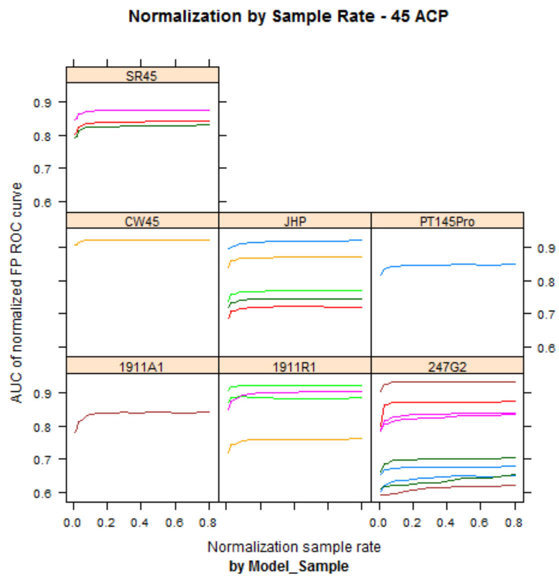

Figure 98: Change in AUC as a function of normalization sample rate for 45 ACP handguns for normalized firing pin scores. A separate plot for each model of handgun is given.

The change in AUC for normalization of the firing pin scores is given in Figure 98. For the Ruger pistols, it can be seen that the relative values for the AUC are higher than those for the breech face scores as given in Figure 99. The wide range of AUCs for the Taurus 24/7 G2 is similar for both firing pin and breech face scores.

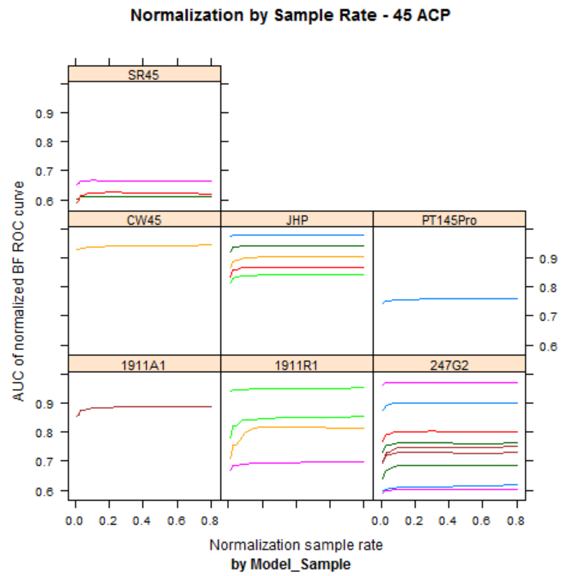

Figure 99: Change in AUC as a function of normalization sample rate for 45 ACP handguns for normalized firing pin scores. A separate plot for each model of handgun is given.

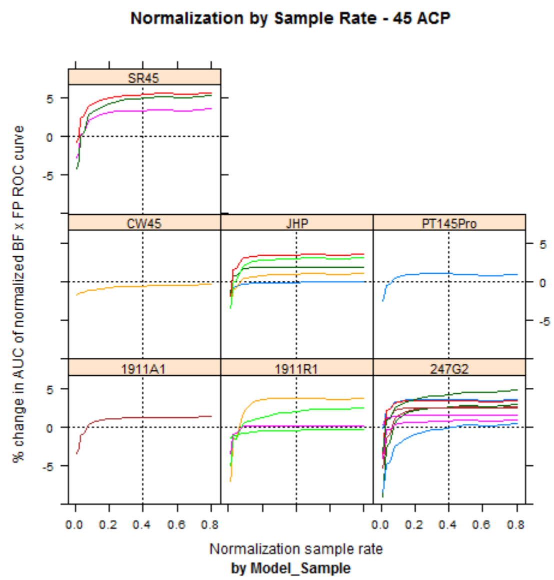

Figure 100: Percentage change in AUC as a function of normalization sample rate for 45 ACP handguns for normalized product of breech face and firing pin scores. A separate plot for each model of handgun is given. The dotted horizontal line indicates no change in AUC and the dotted horizontal line indicates the sample rate (40%) chosen for further analysis

In Figure 100, the product of the breech face score and firing pin scores was normalized and plotted against the sample rate. The product of the scores was obtained before normalization took place. This plot illustrates the raw the rapid initial decrease in AUC at very low sample rates (less than 5%). The rate increases quite dramatically up to about 10%, thereafter there is a gradual increase up to approximately 40%, and thereafter the change is relatively stable. Improvement in the AUC does not exceed 5% for the .45 ACP pistols.

To contrast the changes in Figure 100, the same plot for the .38 Special and .357 Magnum revolvers is given in Figure 101. The changes are quite varied, for instance the Rossi M685 undergoes almost no change in AUC because of normalization. The Ruger New Vaqueros are extremely varied and do not seem to follow a specific pattern. It is also noteworthy that at a 10% sample rate for normalization, some firearms result in a decrease in AUC.

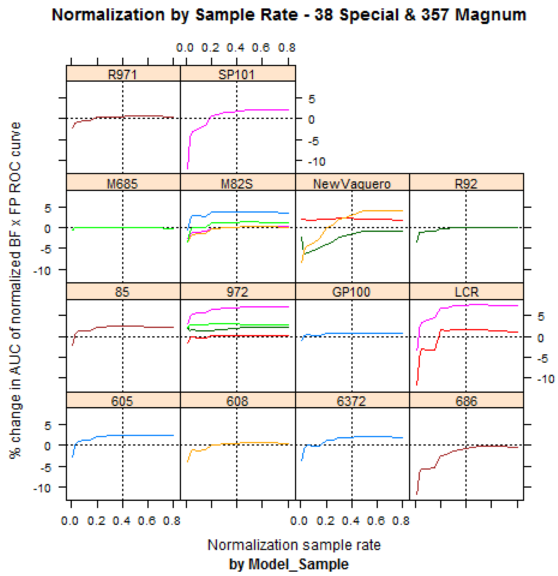

Figure 101: Percentage change in AUC as a function of normalization sample rate for .38 Special and .357 Magnum revolvers for normalized product of breech face and firing pin scores. A separate plot for each model of handgun is given. The dotted horizontal line indicates no change in the AUC value

The percentage improvement of the firing pin score upon normalization was plotted against that for the breech face for individual firearms. The score pairs were marked with the sample rates to assess their influence. The plot (change curve) for a .45 ACP HiPoint JHP Model (X11599) is given in Figure 102. In this plot it is evident that normalization resulted in a reduction of the discriminating ability of both the breech face and firing pin scores. As the sample rate increased, the AUC improved but did not reach the no-change point (0, 0).

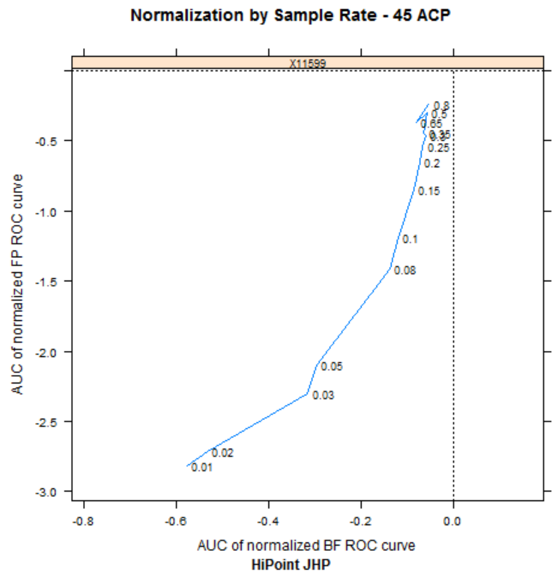

Figure 102: Change curve: HiPoint JHP (.45ACP) (X11599). Percentage improvement in AUC for FP scores as a function of that for the BF scores. The labels are the sample rate for the resultant AUC improvement. The dotted lines indicate the no-change position for the BF and FP AUC.

The change curve for the Taurus PT 145 Pro (X25098) is given in Figure 103. In this example there is a linear relationship between the improvements in AUCs. The breech face score degrades only for the 1% sample rate. The rate of improvement for firing pin scores is greater.

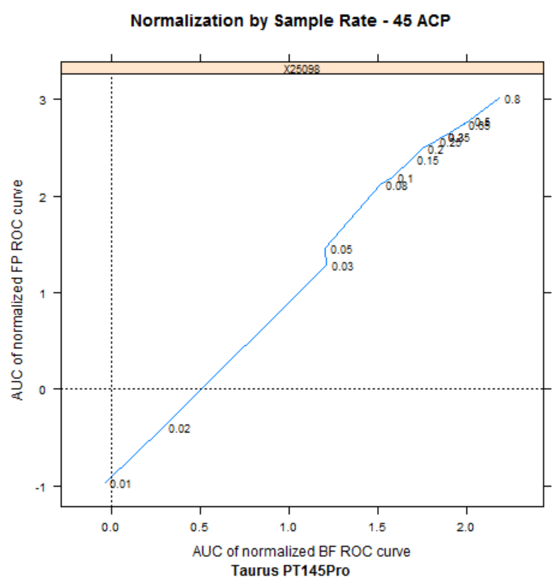

Figure 103: Change curve: Taurus PT 145 Pro (.45ACP) (X25098). Percentage improvement in AUC for FP scores as a function of that for the BF score. The labels are the sample rate for the resultant AUC improvement. The dotted lines indicate the no-change position for the BF and FP AUC.

The change curve for the Taurus 24/7 G2 (XX7938) is given in Figure 104. In this example there is a linear relationship between the improvements in AUCs. The breech face score degrades up to a sample rate of about 6%, thereafter it increases to just less than 1% at a sample rate of 80%. The rate of improvement for the AUCs for the firing pin scores has an immediate increase to 3.2% and then increases slowly to an improvement of about 3.7% at a sample rate of 8%. Thereafter there is a dramatic increase to just less than 8% improvement at an 80% sample rate.

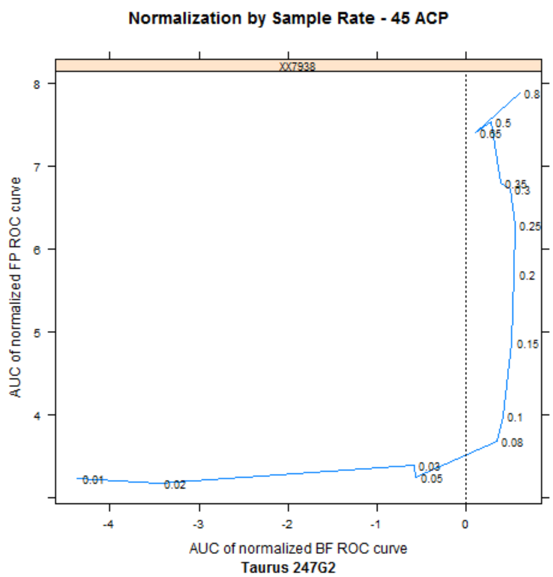

Figure 104: Change curve: Taurus 24/7 G2 (.45ACP) (XX7938). Percentage improvement in AUC for FP scores as a function of that for the BF score. The labels are the sample rate for the resultant AUC improvement. The dotted lines indicate the no-change position for the BF and FP AUC.

The change curve for a Ruger SR45 (X12243) is given in Figure 105. In this example there is a linear relationship between the improvements in AUCs up to a sample rate of approximately 10%. The AUC for the firing pin scores degrade up to a sample rate between 2 and 3%. The improvement maximizes at an improvement of about 2.5% at a sample rate of 80%. The improvement in the AUC for the breech face score is immediately at about 4.9% for a sample rate of 1%. Maximum improvement in the AUC for the normalized breech face score is at a sample rate of about 15%. The AUC improvement degrades by about 1% up to a sample rate of 80%.

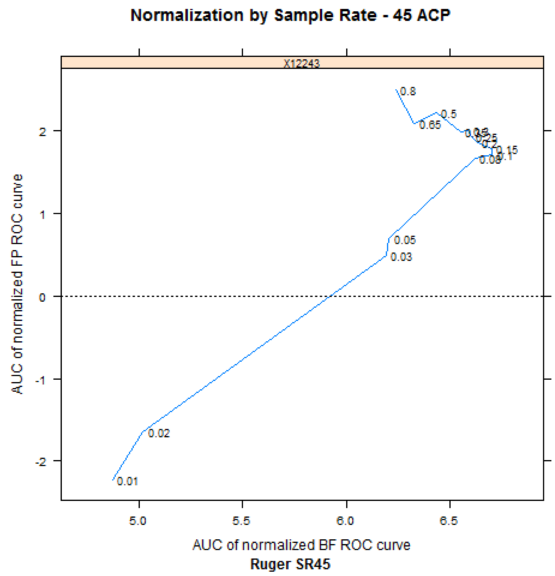

Figure 105: Change curve: Ruger SR45 (.45ACP) (X12243). Percentage improvement in AUC for FP scores as a function of that for the BF score. The labels are the sample rate for the resultant AUC improvement. The dotted lines indicate the no-change position for the BF and FP AUC.

The change curve for a Remington 1911 R1 (X1553A) is given in Figure 106. In this example there is an almost linear relationship between the improvements in AUCs up to a sample rate of approximately 8%. Below this sample rate the AUC for the normalized breech face score has degraded significantly (from a low of about -7%). The AUC for the firing pin scores improve immediately with about 0.25% at a sample rate of 1%. The maximum improvement for firing pin is at the 8% sample rate with an improvement of 2%. Thereafter the AUC degrades for firing pin whilst the gain for breech face start and maximizes at 2% at a sample rate of 80% for normalization.

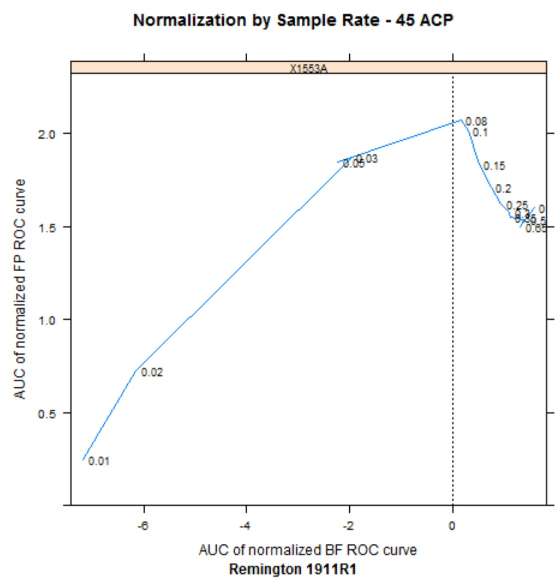

Figure 106: Change curve: Remington 1911 R1 (.45ACP) (X1553A). Percentage improvement in AUC for FP scores as a function of that for the BF score. The labels are the sample rate for the resultant AUC improvement. The dotted lines indicate the no-change position for the BF and FP AUC.

Overall, the previous figures illustrate that for an individual firearm, the normalization percentage is variable, and AUCs may be adversely affected at low normalization rates or at very high rates for a particular firearm. Selection of the specific rate requires a balance of the effects on both the firing pin and breech face scores and is an overall improvement for all firearms. These data suggest that it may be useful to consider both the raw and normalized scores in a single model.

Conclusion

After a discussion with Ultra Electronics Forensic Technology Inc. (the producers of IBIS), a score normalization study was undertaken. Additional derived classifier were introduced, such as FP rank, BF rank, BFxFP rank, BFxFP/BFxFP rank, normalized BF, normalized FP, and normalized BFxFP. As a result of the normalization (at a rate of 10%) there was a small improvement in the raw to normalized area under the receiver operating characteristic curve of 1.76% for BF, 2.07% for FP, and 2.16% for BFxFP. It was also found, using a Remington R1 .45 ACP pistol as an example, that generally the equal error rate improved over the sequence raw score, normalized score, and then rank. In this instance, the order of discrimination was BFFP > BF > FP. Overall it was found that a sampling rate, the proportion of the different-gun score used to determine the mean and standard deviation for the normalization process, of 20% provided the best overall results. Implementation of the normalized system for unknowns proved difficult to implement since the ground truth was unknown and the normalization depends upon knowing which of the candidates represent actual different-source firearms.

Machine-learning of 9 mm data

The aim of this study is to enhance the performance of IBIS system on large-scale databases using the breech face and firing pin correlation scores generated by IBIS. The database used contains pairwise comparisons (6,072,521 pairs) of query cartridges against known firearms entries in IBIS. The database contains cartridges from 9 makes of 9mm caliber firearms (Arcus, Hi-Points, Keltec, Ruger, SCCY, Sig Sauer, Smith & Wesson, Springfield, and Taurus).

To get a measure of the distribution of data, the match and non-match distributions of the entire distribution is plotted in Figure 107 with the breech face (BF) and firing pin (FP) scores as vectors. Normalization is performed by taking the top 10% non-match scores and using them to convert the entire data into their respective *z*-scores.

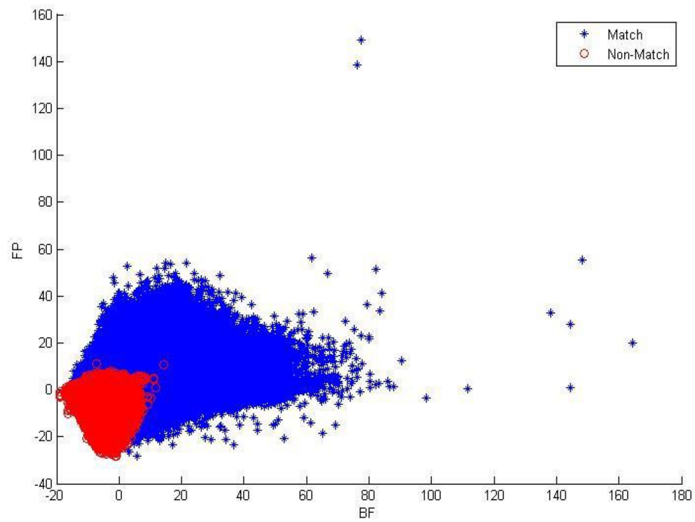

Figure 107: Match and non-match distributions of 9 makes of 9mm caliber firearms with normalized breech face (BF) and firing pin (FP) scores as vectors

To perform the experimental evaluation, the data was first divided in six equal parts, each part containing the same ratio of match and non-match samples. Five of these parts were used for the purpose of 5-fold cross validation and the performances of several machine-learning algorithms are studied. The following machine learning algorithms are summarized below:

- Naïve Bayes: According to Naïve Bayes, assign observation to the most probable class where the assumption is that the features are independent of each other. If $\omega 1$= matching class and $\omega 2$= non-match class for cartridges, then

$$\frac{\mathrm{P}\ (\mathrm{vect}or\ |\ \omega\mathrm{i}\ ) * \mathrm{P}\ (\omega\mathrm{i}\ )}{\mathrm{P}(\omega\mathrm{i}\ \ |\ \mathrm{vect}or)} = (vec\mathrm{tor}),\ \mathrm{i} \in \{0,1\}$$

Where we classify as $\omega1$ if($\omega1$| vector) >P($\omega2$|vector).

- Decision Trees: Decision trees are flowchart-like structures where each node describes an outcome based on particular values of FP and BP (either combined or singular). A node is split recursively based on probability measures. The Gini's Diversity Index (GDI) which measures impurity or node error is used for termination = ($1 - \sum_{i=1}^{M} f_i$) and p is the observed fraction of classes belonging to class *i* reaching a node and R is the total number of classes.
- Bagged Decision Trees: Bagged decision trees are an ensemble of decision trees where many trees are built on dataset sampled from the original dataset with replacement. The logic behind using bagged decision trees is that to reduce the variance and avoid over-fitting in classifier models.
- Neural Networks: In this study, artificial neural networks are used as computational models in supervised learning mode. The BF and FP scores are learned through training and is used to predict a class from unknown data. The nervous system is built by relatively simple units, the neurons**.** They receive and provide information in form of spikes. Simulating functionality of neurons should be able to provide learning ability in algorithms.
- Generalized Linear Model: GLM Creates a response based on a linear function of predictors (FP and BF): $y = 1 + \beta_1 \times BF\ + \beta_2 \times FP$. Most of the time, *y* is assumed to be a normal distribution. Here, a binomial distribution was considered because the response is binary (match or non-match).
- Discriminant Analysis: The main objective of a discriminant function analysis is to predict group membership based on a linear combination of the interval variables. Discriminant analysis creates an equation that will minimize the probability of mislabeling cases into their respective classes.
- KNN: In k-Nearest Neighbors algorithm, an object is labeled by a majority voting of its neighbors. The output class is the most common class among the k nearest neighbors of the object.

Based on the average accuracy of these different machine-learning algorithms, all the data in the five partitions are retrained and the 6th partition is used as a blind test for the best algorithm. The results are summarized in Table 29.

Table 29: Results for cartridge matching using [BF, FP] as feature vector. Accuracy is reported in %

Technique	**Fold 1**	**Fold 2**	**Fold 3**	**Fold 4**	**Fold 5**	**Average**	**Fold 6**	**Fold 6: Match Accuracy**	**Fold 6: Non-match Accuracy**
Naïve Bayes	93.10	93.13	93.09	93.12	93.11	93.11	93.07	49.53	99.45
Decision Trees	90.99	91.00	90.93	91.00	90.96	90.97	91.10	60.39	95.60
Bagged Decision Trees	92.80	92.80	92.77	92.83	92.79	92.79	92.73	59.40	97.62
Neural Networks	93.42	93.48	93.43	93.48	93.45	93.45	93.41	53.37	99.28
GLM	93.37	93.41	93.35	93.40	93.38	93.38	93.34	51.19	99.52
Discriminant Analysis	92.72	92.78	92.70	92.75	92.74	92.73	92.71	43.57	99.91
KNN	90.03	90.07	90.00	90.06	89.97	90.02	90.10	60.61	94.43

Furthermore, the normalized product of BF and FP values (BFFP) was computed and used as a third feature vector and the above experiments were repeated. The results of this test are summarized in Table 30.

Table 30: Results for cartridge matching using [BF, FP, BFFP] as feature vector. Accuracy is reported in %

Technique	**Fold 1**	**Fold 2**	**Fold 3**	**Fold 4**	**Fold 5**	**Average**	**Fold 6**	**Fold 6: Match Accuracy**	**Fold 6 Non-match Accuracy**
Naïve Bayes	93.49	93.55	93.5	93.56	93.5	93.52	93.49	56.46	98.91
Decision Trees	91.09	91.16	91.09	91.18	91.10	91.12	91.25	61.79	95.57
Bagged Decision Trees	93.25	93.28	93.23	93.29	93.26	93.26	93.2	60.42	98.00
Neural Networks	93.42	93.47	93.45	93.47	93.47	93.45	93.43	53.46	99.29
GLM	93.42	93.46	93.41	93.45	93.43	93.43	93.41	51.95	99.48
Discriminant Analysis	92.68	92.73	92.65	92.69	92.69	92.68	92.66	43.04	99.93
KNN	90.38	90.41	90.37	90.41	90.33	90.38	90.44	62.07	94.60

It can be seen that when breech face, firing pin, and the product of these two are used as the feature vector, better discrimination performance is observed compared to using only the breech face and firing pin scores. Due to the overlapping nature of points in matching and non-matching distributions, there is a bias in the results. Most of the points in non-matching distributions are getting correctly classified.

Conclusion

Machine learning of the 9mm data was undertaken using techniques such as naïve Bayes, decision trees, bagged decision trees, neural networks, generalized linear model, discriminant analysis, and k-nearest neighbors. Non-match (different gun) results averaged about 98% whilst match (same-gun) averaged about 54%.

Validation studies

Validation is defined as "the process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended model users"[50]. The aim was to determine if previous studies regarding relationships of IBIS correlation scores, likelihood ratios, and ROC curves could be verified and validated.

Data preparation

The eighteen .40 Smith & Wesson caliber pistols were shot and the cartridge cases were entered into IBIS prior to this particular study. The identifiers, as well as the information about each gun utilized, can be found in Table 31.

---

[50] Smith, Ralph C. "Statistical Validation of Scientific Models." MA 540: Uncertainty Quantification for Physical and Biological Models. North Carolina State University, Spring 2010. Web. 17 Feb. 2015. <http://www4.ncsu.edu/~rsmith/MA797V_S10/Lecture12.pdf>.

Table 31: Identifying Information for all .40 Smith & Wesson firearms used in the validation study

Make	Model	Caliber_Gun	Ammo/Primer_Make	IdentifierGun	String	Match	Type_Sample
Hi-Point	34010	.40 S&W	SST	X71253	HTF		Pistol
Hi-Point	34010	.40 S&W	SCG	X96530	HAF		Pistol
Hi-Point	34010	.40 S&W	SCG	X96531	HBF		Pistol
Hi-Point	34010	.40 S&W	SCG	X96532	HWF		Pistol
Hi-Point	34010	.40 S&W	SCG	X96533	HVF		Pistol
Glock	23 Gen 4	.40 S&W	FA	XMD473	GTF		Pistol
Taurus	24/7 G2	.40 S&W	FA	X34330	TGF		Pistol
Kahr	CW40	.40 S&W	FA	XF0561	KWF		Pistol
Taurus	Millennium Pro 140	.40 S&W	SP	X90724	TMF		Pistol
Smith & Wesson	SD-40 VE	.40 S&W	FA	XE6497	SDF		Pistol
Ruger	SR40	.40 S&W	BB	X41329	RRF		Pistol
Ruger	SR40	.40 S&W	SST	X60581	RVF	RKF	Pistol
Ruger	SR40	.40 S&W	STG	X60581	RKF	RVF	Pistol
Ruger	SR40	.40 S&W	STG	X69508	RGF		Pistol
Springfield	XD40	.40 S&W	FA	X65945	FXF		Pistol
Springfield	XD40	.40 S&W	SST	XX2158	FAF		Pistol
Springfield	XD40	.40 S&W	SST	XX2133	FBF		Pistol
Springfield	XD40	.40 S&W	SST	XX2135	FGF		Pistol

Each of the *.csv files contain all of the information necessary to read the *.txt files. An example of the first few lines of a DataFile is given in Table 32.

Table 32: DataFile example for FAF, a Springfield XD40.

GunFile	AmmoFile	DateFile	SeqFile
FAF	UK-SST	713	301
FAF	UK-SST	713	302
FAF	UK-SST	713	303

Along with the DataFiles folder, there are also folders for each string, a BFR (breech face rank) folder, and FinalClean folder (Z:\40 S&W) for output.

The .txt files are cleaned using the cleanFiles_40S&W_.R script. The data in the clean .csv files are then ranked by breech face using the Breech face Rank 40S&W.R and a "final.csv" is created. These final files are combined using a batch file created in Notepad++[51]. These files are combined to analyze and assess the caliber as a whole, instead of each individual gun.

Validation procedure

An Excel® file was created to track progress of validation, which can be found in Table 33. This table shows the major steps in the validation process: the creation of the files by splitting the data frames, creating the Bayesian networks (BN) and processing the data in these networks, and generating the final Excel® data sheet including the output data generated from the processing of the Bayesian networks.

Table 33: Bayesian network progress tracker

		Files Created		BN Created		Final Excel		Complete
Sampling	**Seed**	*Raw*	*Norm*	*Raw*	*Norm*	*Raw*	*Norm*	
0.1	42							
	84							
	168							
	336							
	672							
0.2	42							
	84							
	168							
	336							
	672							
0.3	42							
	84							
	168							
	336							
	672							

Files created - splitting data frames

The final file (Z:\Files - Final Clean\40S&W\40SWFinal) is then split into two separate data frames: training and testing. This is done using the Split Data Frame into training and testing 40S&W.R script. Each line of coding is shown below for the Raw score files, following an explanation of its function. The Raw score files consists of the columns BF, FP, and BFxFP. (These columns are different from what the Normalized score files utilize, which will be discussed shortly).[52]

1. C40SW <- read.csv(Z:/Files - Final Clean/40S&W/40SWFinal.csv")
2. #If you still have file headers present, remove and save the file.

[51] https://notepad-plus-plus.org/

[52] The lines beginning with "#" are meant to be notes within the script for others to be able to utilize and follow along.

3. CC40SW <- subset(CC40SW, Rank!="Rank", select=c(CaseID_Sample, ExhibitNumber_Sample, Rank, CaseID_DB, ExhibitNumber_DB, BF, FP, Match, Make_DB, Model_DB, Ammo_DB, Caliber_DB, Firing_Pin_Type_DB, Make_Sample, Model_Sample, Ammo_Sample, IdentifierGun_Sample, IdentifierGun_DB, Caliber_Sample, Firing_Pin_Type_Sample, Type_Sample, Primer_Sample, Primer_DB, Drag_Mark_Sample, Drag_Mark_DB, Reload, Rank_BF, BFFP, CaseID_pre, BF_norm, FP_norm, BFFP_norm, Same_Model))
4. write.csv(CC40SW, "Z:/Files - FinalClean/40SWFinal.csv",row.names=FALSE)

The prepared file is read into memory. When the files were made in Excel®, column headers were left intact. These column headers are then removed. The header-less file is then saved.

5. #create an R object for the data...will load MUCH faster (file is almost 15X smaller)
6. save(CC40SW, file = "Z:/Files - Final Clean/40S&W/40SWFinal.RData")
7. load(file = "Z:/Files - Final Clean/40S&W/40SWFinal.RData")
8. #training = 90% - testing = 10%
9. rate <- 0.1

Three rates were used: 0.1, 0.2, and 0.3. The data is split into two different frames: testing and training. The rate in line nine indicates the testing data frame size. In other words, if the rate is 0.1 (equating to 10%), then the training frame would contain the other 90% of the data. The reason the data is split is to be able to determine if the method is working, to validate itself against the test set.

10. samplesize <- ceiling(length(CC40SW$Rank)*rate)
11. set.seed(42) # change the seed when you run a new evaluation.

Five seeds were used: 42, 84, 168, 336, and 672. This ensures that the random splitting of the file is different in each instance, but since specific seeds are used the process can be repeated. The entire evaluation combinations can be seen in Table 34.

Table 34: Evaluation Sampling and Seeding Rates.

Sampling	Seed				
0.1	42	84	168	336	672
0.2	42	84	168	336	672
0.3	42	84	168	336	672

12. #Raw Scores
13. CC40SW <- subset(CC40SW, select=c(Rank, BF, FP, Match, Make_DB, Model_DB, Make_Sample, Model_Sample, IdentifierGun_Sample, IdentifierGun_DB, Drag_Mark_Sample, Drag_Mark_DB, Rank_BF, Same_Model))
14. test_set_data <- sample(length(CC40SW$Rank), size = samplesize, replace = FALSE)
15. CC40SW_test <- CC40SW[test_set_data,]
16. CC40SW_train <- CC40SW[-test_set_data,]

The final test and training files are created using code lines thirteen through sixteen. The columns of interest are highlighted in line thirteen. Not all of the columns generated in the original files were needed; therefore they were eliminated in order to allow for quicker file processing. The test data set was created using the sampling and seeding rates, such as 0.1 and 42. The training data set was essentially the counterpart to the test data set.

```
17. #File for use in Netica must be in .csv format
18. write.csv(CC40SW_test, "Z:/Firearms/Firearms/Files – Final
    Clean/40S&W/CC40SWtest_0.1&42.csv",row.names=FALSE)
19. write.csv(CC40SW_train, "Z:/Firearms/Firearms/Files - Final
    Clean/40S&W/CC40SWtrain_0.1&42.csv",row.names=FALSE)
```

The final lines of code, eighteen and nineteen, save the files as a comma delimited version (*.csv) of the Excel® file. Netica™, the software used to create the Bayesian networks, require the files to be in this mode to be processed.

The entire process explained above is then repeated with the columns of the files that contain the normalized data. Normalization originates from statistics and eliminates the unit of measurement by transforming the data into new scores (z-scores) with a mean of zero and a standard deviation of one. Normalizing a set of scores involves subtracting the sample mean from the score and then dividing by the standard deviation of the sample. For the purpose of this research, the mean and standard deviation of a variety of sampling percentages of non-match scores for each firearm was found and then used to convert each cartridge case fired from that firearm to a z -score. This was performed for firing pin, breech face, and their product.

The normalized data was created when the Breech face Rank 40S&W.R script was run. The normalized data consists of the following columns: BF_Norm, FP_Norm, BFFP_Norm. The changes to the script above includes inserting "Norm" into the file names and using the following code instead of what is used in line thirteen:

```
12. #Normalized Scores
13. CC40SW <- subset(CC40SW, select=c(Rank, Match, Make_DB, Model_DB, Make_Sample,
    Model_Sample, IdentifierGun_Sample, IdentifierGun_DB, Drag_Mark_Sample, Drag_Mark_DB,
    Rank_BF, BF_norm, FP_norm, BFFP_norm, Same_Model))
```

The raw data columns have been replaced with the columns that are important to process the normalized data.

Creating Bayesian networks

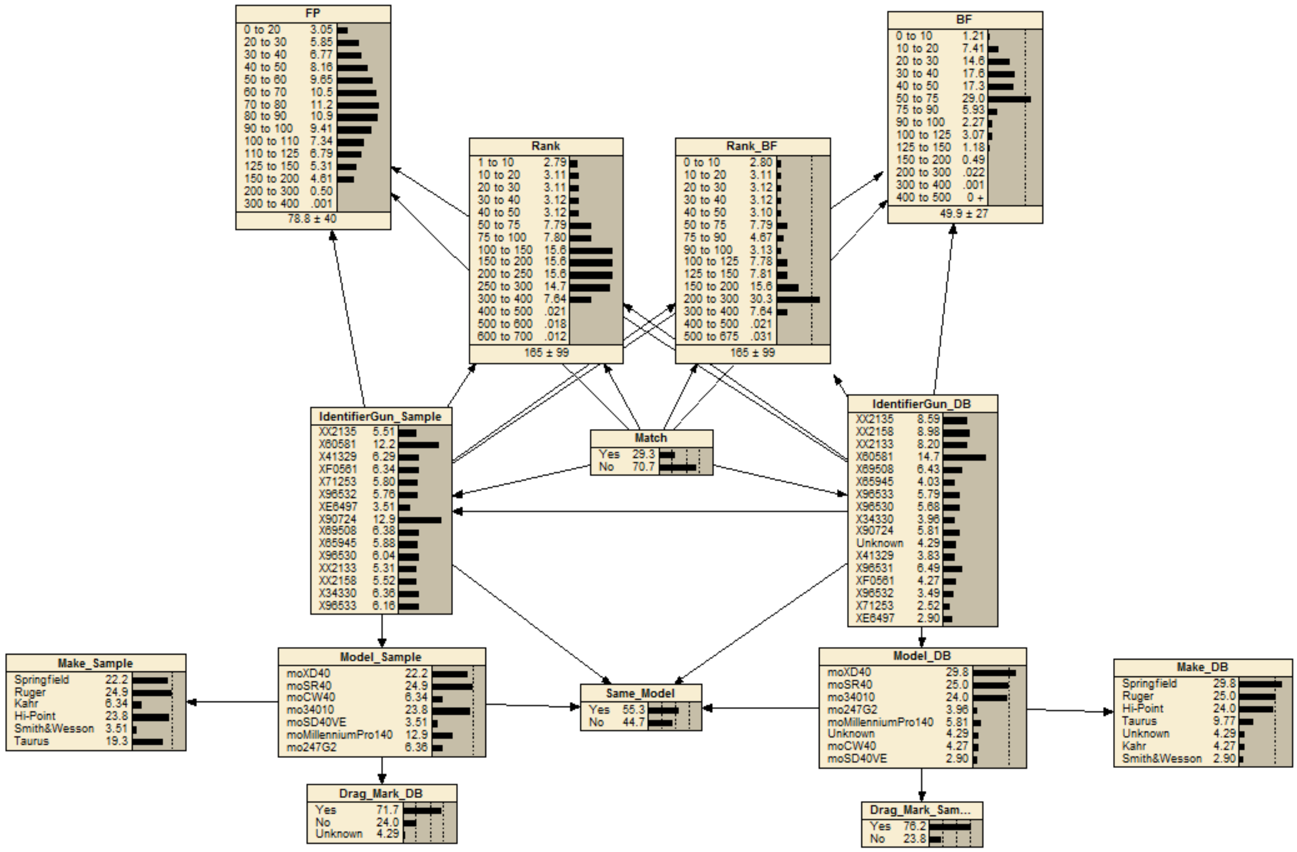

Figure 108: Bayesian network used for validation (sampling rate = 0.1 & seed = 42)

The Bayesian networks, an example shown in Figure 108, were created using Netica™, a Norsys™ Software Corp application. The base of the network was created by reading in a "new case", *i.e.* the training data file created above. This network was then learned. Norsys™ describes learning as "the process of automatically determining a representative Bayes net given data in the form of cases (called training cases). Each case represents an example, event, object, or situation in the world and the case supplies values for a set of variables which describes the event." The completion of case learning initiated the second step: case file processing. To process a case, Netica™ requires two files: the control file and the test file. The control file is used to generate the columns of the output file, which contains the posterior-likelihood beliefs. The control file created for the validation process utilized beliefs. The control file can be found in Figure 109.

```
bel(Match, Yes)
bel(Match, No)
```

Figure 109: Control file used to process case files in Netica™

The “Match” column in the Excel® files were used for validation because it was determined this would be the most appropriate contributing factor to reinforce that a proper method was being utilized. The control file above displays the belief that the value of match being either yes or no will be the basis for the probabilities produced. The second important file is the corresponding test file for the training file selected as the case that was previously learned. The completion of case processing produces the output file. An example of a few lines from the output file can be seen in Figure 110, where P(+Match) equals the probability of a match given the evidence (P(Match = Yes|E) and likewise with P(-Match) equaling the probability of a match given the evidence (P(Match = No|E).

```
P(+Match)    P(-Match)
0.00215506   0.997845
1   1.14956e-09
0.00898683   0.991013
0.166733     0.833267
```

Figure 110: Sample of output file generated from casefile processing

Importing output data from Bayesian network into Excel® - final Excel®

The output beliefs generated above were added to the test files.

Equation 6: Calculation of likelihood ratios using the posterior odds and prior odds generated in Netica™

$$\frac{PosteriorMatchYes * PriorMatchNo}{PosteriorMatchNo * PriorMatchYes} = LR$$

The likelihood ratio (LR) was then calculated using the posterior-odds (Posterior_Match_Yes, Posterior_Match_No) from the output file and the prior-odds (Prior_Match_Yes, Prior_Match_No) generated from case learning. The log likelihood ratios (LLR) were also calculated.

Equation 7: Calculation of the log(likelihood ratio)

$$\log(LR) = LLR$$

Data analysis

The completed Excel® files were analyzed further by creating ROC curves and area under the curve. The ROC curves were also used to generate the error rate curves. The ROC curve demonstrates the discriminating power of the method. This discriminating ability is directly related to the area under the ROC curve. The error associated with this method is determined by the parameter under evaluation. The *fpr* and the *fnr a*re given as a function of the correlation scores that were obtained by the IBIS. The crossover from black to gray to white zones are indicated when the error rates are zero. The grey is where the match and non-match scores overlap. It is in this gray zone where the quality-quantity relationship is the most critical. Two tables were generated to show the AUC values, along with the averages and standard deviations for each sample rate.

Table 35: Area under the curve values using the LLR developed from the raw scores for all 40S&W pistols

RAW				
Sample Rate	**Seed Value**	**AUC**		
0.1	42	0.909		
	84	0.909	Average	Standard Deviation
	168	0.911	91.0%	0.16%
	336	0.908		
	672	0.911		
0.2	42	0.910		
	84	0.909	Average	Standard Deviation
	168	0.911	91.0%	0.08%
	336	0.909		
	672	0.910		
0.3	42	0.910		
	84	0.909	Average	Standard Deviation
	168	0.910	91.0%	0.05%
	336	0.909		
	672	0.910		

The raw scores were utilized to calculate the data in Table 35. As the sample rate gets higher in percentage (i.e. from 10% to 30%), the standard deviation becomes smaller. This analysis provides evidence that as the sample rate increases, the data better approximate the true value.

Table 36: Area under the curve values using the LLR developed from the normalized scores for all 40S&W pistols

NORM				
Sample Rate	**Seed Value**	**AUC**		
0.1	42	0.908		
	84	0.892	Average	Standard Deviation
	168	0.899	90.2%	0.66%
	336	0.906		
	672	0.902		
0.2	42	0.900		
	84	0.901	Average	Standard Deviation
	168	0.903	90.3%	0.17%
	336	0.904		
	672	0.904		
0.3	42	0.904		
	84	0.904	Average	Standard Deviation
	168	0.905	90.5%	0.08%
	336	0.905		
	672	0.906		

The normalized scores were utilized to calculate the data in Table 36. Similar to Table 35, the standard deviation decreases as the sample rate increases.

These calculations show slight change; however, there was no significant difference from sample rate to sample rate. This indicates that the data, as a caliber, is tightly gathered around the mean and thus more reliable because there is little variation.

Conclusion

A validation study of the proposed Bayesian network was undertaken by subdividing the data into test and training sets using random selection of samples. The test sets were run and evaluated by the ability of the network to correctly classify the sample. The averages of the areas under the curve were about 91% with a standard deviation of less than 0.2%, which decreased as the sample size increased.

2D and 3D IBIS study

The cartridge cases from a sample set of twelve 9 mm firearms were used to study 3D correlations with cooperation of Ultra Electronics Forensic Technology Inc. A breakdown of the identifying components of each firearm can be found in Table 37.

Table 37: Identifying information of twelve 9mm firearms compromising the 2D/3D study sample set

Make	Model	Identifier
SCCY	CPX	X66727
Springfield	XD9	X17802
Keltec	P11	XAZV54
Ruger	LC9	X43521
Springfield	XD9	X77862
Keltec	Sub2000	XEF603
HiPoint	C9	X55426
Arcus	D98	XXX724
SCCY	CPX	X97571
Taurus	247G2	X45405
Keltec	PF9	XSBP59
Ruger	SR9	X69363

These twelve firearms were selected based on preliminary data which displayed their performances of breech face (BF) and firing pin (FP) IBIS scores via their receiver operating characteristic (ROC) curves and the accompanying area under the curve (AUC) values. ROC curves can be used to determine the crossovers between match and non-match. The ROC curve demonstrates the discriminating power of the method. In other words, it determines how well the method can differentiate between different states of the samples to which the method has been applied. This discriminating ability is directly related to the area under the ROC curve. Figure 111 displays the AUC scores of the sample set of firearms by make and model.

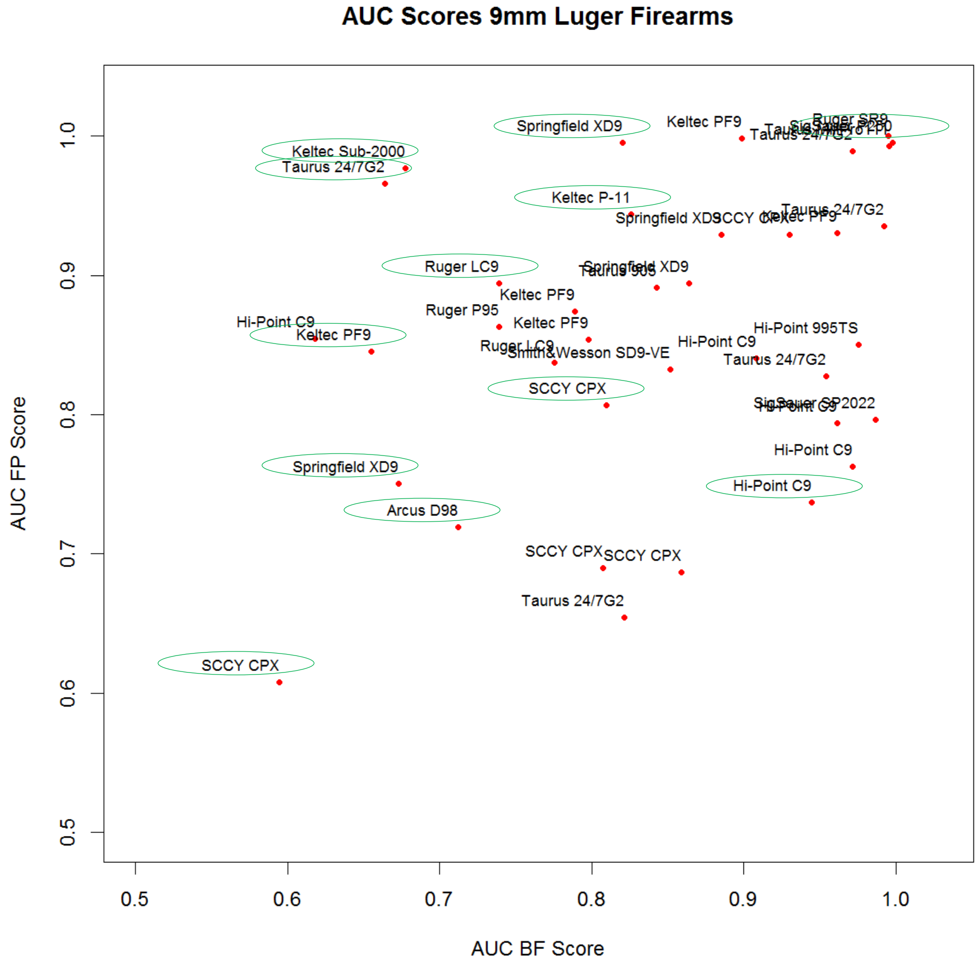

Figure 111: AUC scores of 9mm Luger Firearms using the product of BF and FP correlation scores of IBIS

The firearms circled in green indicate the firearms that were chosen to take to Ultra Electronics Forensic Technology Inc. to perform analysis using their 3D instrument. These firearms were selected as an appropriate representation of the product of BF and FP AUC score performances of 9 mm Luger firearms within the West Virginia University database. The goal of this study was to perform a 3D IBIS analysis and compare these results to that of a 2D IBIS analysis. The intra- and inter-variability of those scores for 9mm Luger firearms was also analyzed. This explains why some makes and models were selected more than once to comprise the sample set.

These cartridge cases were taken to FTI headquarters and analyzed. The correlation scores were printed, converted to Excel® files, and then run through an R script in RStudio to produce ROC curves and AUC values to visually display the data. In order to better analyze the data, it was broken down by different filters: firearm, category, and instrument. The complete list of AUC scores for each category, firearm, and instrument can be found in Table 38 (FTI) and Table 39

(WVU). The FTI IBIS has capabilities of analyzing firearms evidence in 2D and 3D, whereas the WVU IBIS can only analyze in 2D.

Table 38: All AUC scores of 2D and 3D data collected from FTI

Make Sample	Model Sample	GunID Sample	auc.BF 2D_SideLight	auc.FP_3D	auc.FP_2D	auc.BF_3D	auc.BF_2D
SCCY	CPX	X66727	1.000	0.924	0.978	0.578	0.870
Springfield	XD9	X17802	0.728	0.994	0.999	0.917	0.668
Keltec	P11	XAZV54	0.924	0.980	0.974	0.896	0.781
Ruger	LC9	X43521	0.678	1.000	1.000	0.738	0.506
Springfield	XD9	X77862	0.680	0.999	0.982	0.854	0.583
Keltec	Sub2000	XEF603	0.933	0.997	0.995	0.989	0.772
HiPoint	C9	X55426	1.000	0.992	0.964	0.998	0.912
Arcus	D98	XXX724	0.830	0.983	0.993	0.890	0.873
SCCY	CPX	X97571	0.996	0.735	0.884	0.876	0.828
Taurus	247G2	X45405	0.846	0.981	0.973	0.624	0.609
Keltec	PF9	XSBP59	0.976	0.987	0.923	0.893	0.759
Ruger	SR9	X69363	0.998	1.000	0.999	0.922	0.936

The worst discriminating power category from Ultra Electronics Forensic Technology Inc. with respect to all the firearms analyzed is BF_2D whereas the best discriminating power category is FP_3D. Also found in Table 39 are the normalized scores from the WVU data.

Table 39: All AUC scores of 2D data collected from WVU (norm=normalized)

Make Sample	Make Model	GunID Sample	BF	FP	FPBF	Rank FP	Rank BF	Rank FPBF	Score Over Rank	BF (norm)	FP (norm)	FPBF (norm)
SCCY	CPX 2	X66727	0.656	0.989	0.926	0.993	0.684	0.926	0.923	0.676	0.992	0.932
Springfield	XD9	X17802	0.548	**1.000**	0.978	0.999	0.554	0.978	0.982	0.547	1.000	0.982
Keltec	P11	XAZV54	0.684	0.986	0.937	**0.990**	0.736	0.937	0.945	0.729	0.988	0.954
Ruger	LC9	X43521	*0.447*	**1.000**	0.958	**1.000**	*0.470*	0.958	**0.987**	*0.461*	**1.000**	0.982
Springfield	XD9	X77862	0.666	0.990	0.970	0.978	0.663	0.970	0.953	0.649	0.978	0.951
Keltec	Sub 2000	XEF603	0.513	**0.996**	0.957	0.995	0.530	0.957	0.959	0.508	0.995	0.951
HiPoint	C9	X55426	0.890	0.962	0.972	*0.962*	**0.930**	0.972	0.978	**0.928**	*0.964*	0.982
Arcus	D98	XXX724	0.808	0.997	**0.987**	0.997	0.829	**0.987**	0.984	0.823	**0.997**	0.987
SCCY	CPX	X97571	0.713	0.976	0.963	0.965	0.723	0.963	0.950	0.703	0.968	0.961
Taurus	24/7 G2	X45405	0.531	0.977	*0.901*	0.973	0.553	*0.901*	*0.901*	0.545	0.972	0.898
Keltec	PF9	XSBP59	0.725	*0.960*	0.950	0.966	0.732	0.950	0.973	0.721	0.965	0.939
Ruger	SR9	X69363	**0.911**	0.972	0.973	0.970	0.910	0.973	0.962	0.905	0.969	0.970

The worst discriminating power category from WVU with respect to all the firearms analyzed is the BF scores while the best is the FP scores.

Table 40 displays the minimum and maximum scores, and their associated categories, filtered by firearm.

Table 40: 2D and 3D data observations sorted by firearm displaying the category and the AUC score

Data Observations by Firearm

			Minimum Score		**Maximum Score**	
Make	**Model**	**Identifier**	**Metric (auc)**	**Score**	**Metric (auc)**	**Score**
SCCY	CPX	X66727	BF 3D	0.578	BF 2D SideLight	1.000
Springfield	XD9	X17802	BF (norm)	0.547	FP (norm)	1.000
Keltec	P11	XAZV54	BF	0.684	Rank FP	0.990
Ruger	LC9	X43521	BF	0.447	FP, Rank FP, FP (norm), FP 3D, FP 2D	1.000
Springfield	XD9	X77862	BF 2D	0.583	FP 3D	0.999
Keltec	Sub2000	XEF603	BF (norm)	0.508	FP 3D	0.997
HiPoint	C9	X55426	BF	0.890	BF 2D SideLight	1.000
Arcus	D98	XXX724	BF	0.808	FP (norm)	0.997
SCCY	CPX	X97571	BF (norm)	0.703	BF 2D SideLight	0.996
Taurus	247G2	X45405	BF	0.531	FP 3D	0.981
Keltec	PF9	XSBP59	BF (norm)	0.721	FP 3D	0.987
Ruger	SR9	X69363	BF (norm)	0.905	FP 3D	1.000

The data was separated by firearm in order to analyze the intra- and inter-variability between the same makes as well as the same models with different identifiers (serial numbers). The SCCY CPX 2 firearms performed the best with regards to 2D BF scores; however, they did not perform the same and have two separate maximums and minimums. This observation indicates that BF score has the best discriminatory power for SCCY CPX II firearms. The Springfield XD9 firearms performed highest with regards to FP scores and lowest with BF scores, indicating that FP has the better discriminatory power. Similar to that of the SCCYs, these two firearms of same make and model did not perform the same. There were three Keltec firearms analyzed of three different models: P11, Sub-2000, and PF9. All three performed the best with respect to FP scores and the worst with BF scores indicating a class characteristic that the FP has a higher discriminatory power than the BF. The Sub-2000 and the PF9 performed similarly both having auc.FP_3D as the highest score and auc.BF_norm as the lowest score, whereas the P11 had the highest value with auc.RankFP and the lowest with auc.BF. The two Ruger firearms, LC9 and SR9, performed similarly in the fashion that the FP had the highest scores and the BF had the lowest. The LC9 performed the same across five categories of FP scores resulting in a value of

1. The only HiPoint performed best using the 2D Sidelight feature of BF analysis and the worst at the standard BF position. Unlike the other makes, it is unclear if BF or FP is a more discriminatory feature of a cartridge case from a HiPoint firearm. The Arcus D98 and the Taurus 24/7 G2 can be better identified from the FP impression than from the BF, which is reflected in their minimum and maximum scores. Overall, with respect to all the firearms examined, every minimum value is derived from the BF scores (2D, 3D, or normalized).

Similarly to Table 40, Table 41 displays the minimum and maximum scores, and their associated firearm, filtered by category. While some of the data is a repeat from the tables above, it provides a different analysis perspective based on the categories.

Table 41: 2D and 3D data observations sorted by category displaying the firearm's information and AUC score

Data Observations By Category

Category (auc)	**Minimum Score**	**Maxiumum Score**
BF	Ruger LC9 X43521 (0.447)	Ruger SR9 (X69363) (0.911)
FP	Keltec PF9 XSBP59 (0.960)	Ruger LC9 X43521 (1.000)
FPBF	Taurus 247G2 X45405 (0.901)	Arcus D98 XXX724 (0.987)
Rank FP	HiPoint C9 X55426 (0.962)	Ruger LC9 X43521 (1.000)
Rank BF	Ruger LC9 X43521 (0.470)	HiPoint C9 X55426 (0.930)
Rank FPBF	Taurus 247G2 X45405 (0.901)	Arcus D98 XXX724 (0.987)
Score Over Rank	Taurus 247G2 X45405 (0.901)	Ruger LC9 X43521 (0.987)
BF (norm)	Ruger LC9 X43521 (0.461)	HiPoint C9 X55426 (0.928)
FP (norm)	HiPoint C9 X55426 (0.964)	Ruger LC9 X43521 (1.000)
FPBF (norm)	Taurus 24/7G2 X45405 (0.898)	Arcus D98 XXX724 (0.987)
BF 2D SideLight	Ruger LC9 X43521 (0.678)	HiPoint C9 X55426, and SCCY CPX X66727 (1.000)
FP 3D	SCCY CPX X97571 (0.735)	Ruger LC9 X43521, and Ruger SR9 X69363 (1.000)
FP 2D	SCCY CPX X97571 (0.884)	Ruger LC9 X43521 (1.000)
BF 3D	SCCY CPX X97571 (0.577)	HiPoint C9 X55426 (0.998)
BF 2D	Ruger LC9 X43521 (0.506)	Ruger SR9 X69363 (0.936)

For both systems, the Ruger LC9 (X43521) had the lowest value resulting from the BF scores (0.447 WVU and 0.506 for FTI). Figure 112 illustrates the poor performance of the Ruger LC9 via a scatterplot of the BF scores. It is interesting to note that the Ruger LC9 was the worst performance in both 2D BF categories while the Ruger SR9 performed the highest.

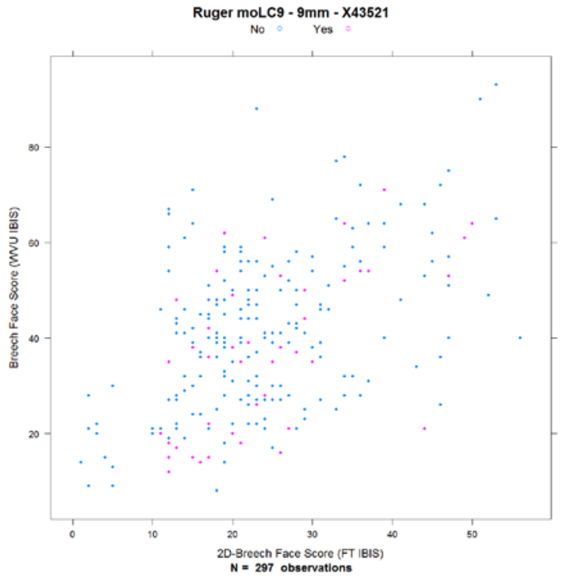

Figure 112: Scatterplot comparing the 2D BF match and non-match scores from FTI and WVU of a Ruger LC9 (X43521.

In Figure 112, the blue dots represent a non-match while the pink represent a match. The lack of a clear separation of distribution, along with some of the non-match scores being higher than the match scores, accounts for the poor performance and low scores from this region of interest. A possible explanation could be a privation of discriminatory impressions made from the BF of a Ruger LC9. In contrast, Figure 113 shows the superior performance of the Ruger SR9 via a scatterplot of the BF scores.

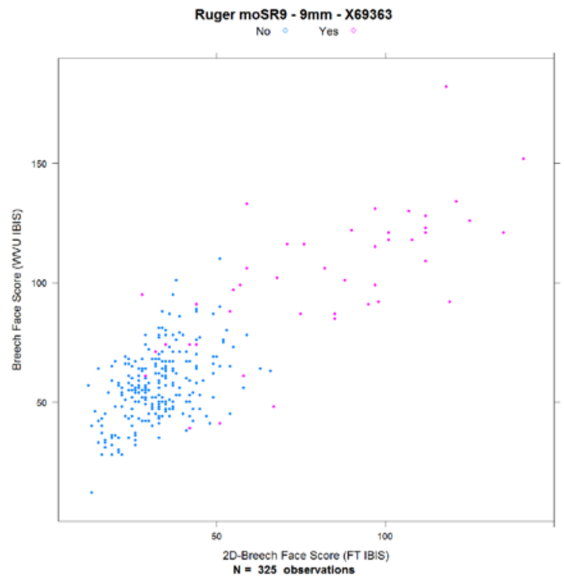

Figure 113: Scatterplot comparing the 2D BF match and non-match scores from FTI and WVU of a Ruger SR9 (X69363)

Again, the blue dots represent a non-match while the pink represent a match. There is a clear separation between the distributions of scores, and as expected, the match scores are higher than the non-match scores. This scatterplot shows that the BF of a Ruger SR9 has a high discriminatory power. One conclusion that can be made from the comparison of Figure 112 and Figure 113 is that the quality of performance of the BF impressions is not the same across different models of Ruger firearms. If the analysis of the SR9 had not been included in this study, one might assume that poor performance of BF scores is a class characteristic of all 9mm Ruger firearms.

In the category of 2D and 3D FP, the performance of the Ruger LC9 is the best, with an AUC of 1.000. Figure 114 displays the improved performance of the Ruger LC9 with firing pin scores.

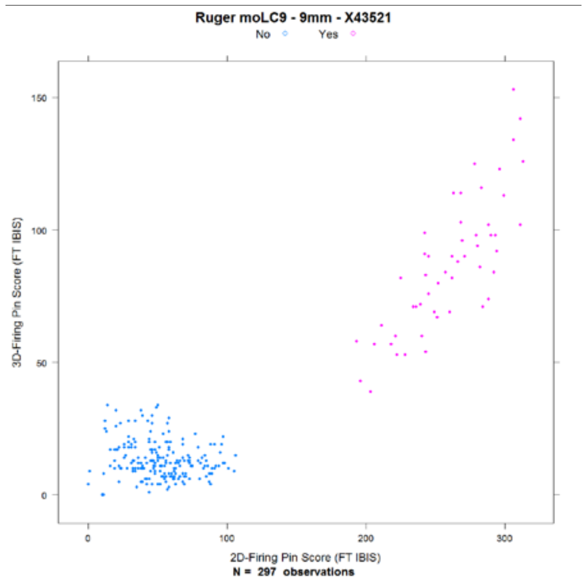

Figure 114: Scatterplot comparing the 3D and 2D FP match and non-match scores from FTI of a Ruger LC9 (X43521)

There is a clear separation on both axes. The overall match scores are higher than the non-match scores, as expected. This scatterplot shows that the FP of a Ruger LC9 has a high discriminatory power. Another part of this study was to determine if there is a significant difference in using an instrument with 3D capabilities versus one with 2D capabilities. The scores of the 2D FTI IBIS were not significantly different from those of the 2D WVU IBIS making them comparable. Figure 114 also shows that even using 3D technology, the FP score is still highly discriminatory and shows clear separation in its distribution of match and non-match scores. The performance of the Ruger LC9 is highly variable: it has both the lowest and highest scores across more categories than any other firearm, the worst in BF and the best in FP. Figure 115 displays the density distributions of the scores for FP, BF, and their product (BFxFP) for the Ruger LC9.

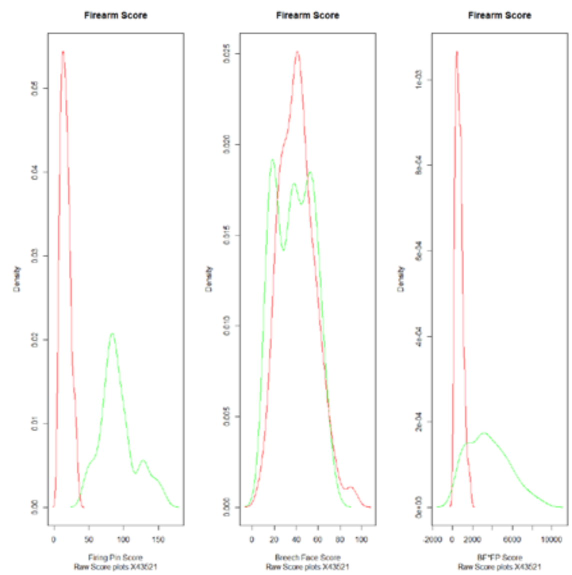

Figure 115: Density distributions for the FP, BF, and BFFP (product) scores for the Ruger LC9 (X43521) obtained from the WVU IBIS

The red curve (Figure 115) represents the non-match score distribution while the green curve represents the match score distribution. In both the FP score and product score distributions, there is clear separation indicating a high discriminatory value for firearms analysis. In comparison with Figure 112, the BF score distribution shows a lack of separation indicating a low discriminatory value for firearms identification. The case of the FP scores being significantly higher than the BF scores was not the case for all firearms, as it was for the Ruger LC9 (X43521). The Arcus D98 (XXX724) had higher values (not by much) for the FP scores than for the BF scores, but there was still clear separation between the two impression areas. Figure 116 shows the distribution densities of the FP, BF, and product scores for the Arcus D98 (XXX724).

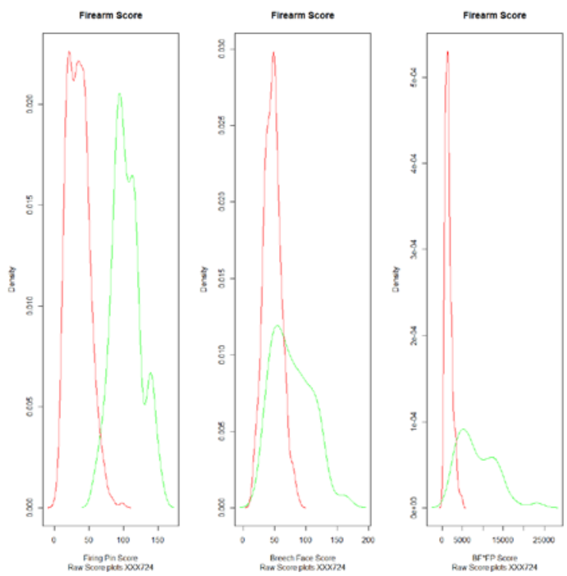

Figure 116: Density distributions for the FP, BF, and BFFP (product) scores for the Arcus D98 (XXX724) obtained from the WVU IBIS

The distributions are clearly separated for each category; however, there is an overlap in the BF density plot, which correlates with its lower scores.

There is a lack of significant performance, with regard to BF and FP, for the Springfield XD9s when there was a clear separation between the two from Figure 111. Also according to Figure 111, the SCCY CPX II (X66727) should have been the worst performer; however, it in fact performs as one of the best in the 2D side light feature of BF. This can be explained by Figure 111 data being from WVU while the side light feature comes from the FTI instrument. In a ROC curve the true positive rate (sensitivity) is plotted as a function of the false positive rate (specificity). Each point on the ROC curve represents a sensitivity/specificity pair corresponding to a particular decision threshold. A test with perfect discrimination, no overlap of the two distributions, has a ROC curve that passes through the upper left corner (100% sensitivity, 100% specificity). Therefore, the closer the ROC curve is to the upper left corner, the higher the overall accuracy of the test[53]. A perfect ROC curve can be found below in Figure 117.

[53]Zweig MH, Campbell G (1993) Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clinical Chemistry 39:561-577.

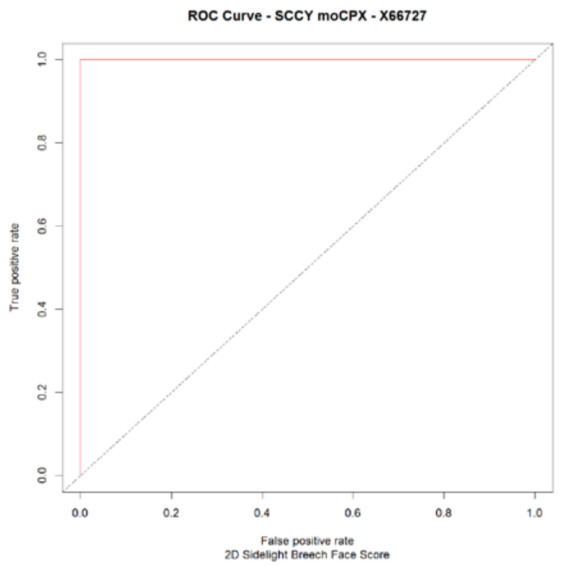

Figure 117: 2D side light BF ROC curve for the SCCY CPX II (X66727)

The value for the AUC of the ROC curve in Figure 117 is 1.000, indicating perfect discrimination. Other performances with perfect discrimination (ROC curve identical to Figure 74 and AUC equal to 1.000) are the HiPoint C9 (X55426) in the 2D side light feature of BF and the Ruger LC9 (X43521) for all categories with FP solely (2D, 3D, norm and rank). Conversely, the other SCCY CPX II (X97571) appears as the lowest scores for 2D FP and 3D FP and BF resulting in a ROC curve similar to that of the Taurus 24/7 G2 (X45405) found in Figure 118.

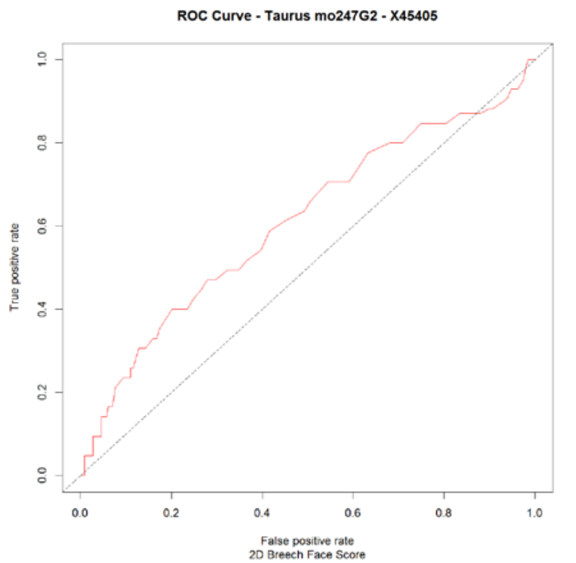

Figure 118: 2D BF ROC curve for the Taurus 24/7 G2 (X45405)

ROC curves such as that in Figure 118 are far from the desired upper left corner. This indicates a poor performance with regards to accuracy (i.e. ability to discriminate between the two distributions). The dotted line in the middle represents 50% specificity and 50% sensitivity, making the distribution of match and non-match for 2D BF scores no better than a coin flip.

Comparing all of the data from both instruments, they behaved similarly resulting in the worst performance resonating from a Ruger LC9 in the category of 2D BF scores. Also noteworthy is the benefit of the addition of the side light feature for analyzing the BF. Overall, with regards to an added dimension (*i.e.* 2D vs 3D), there was no significant difference in the results to conclude that one system is better than the other.

A general comparison of performance of the two systems is given in Figure 76 (breech face) and Figure 77 (firing pin). The linear regression (solid line) and the $y = x$ (dashed line) indicates the similarity in scores. The variability is assigned to user and sample orientation. From Figure 119 it appears as if the breech face match scores follow the $y = x$ line and the regression is weighted to the non-match scores. In Figure 120, the firing pin match scores follow the regression line, but at higher scores the FT system deviates to higher values than the WVU system.

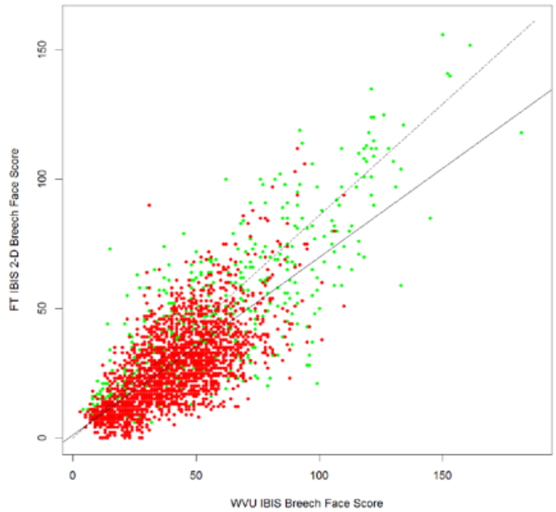

Figure 119: Comparison of breech face scores of the WVU Legacy IBIS and the Forensic Technologies 3D IBIS using the 2D scores

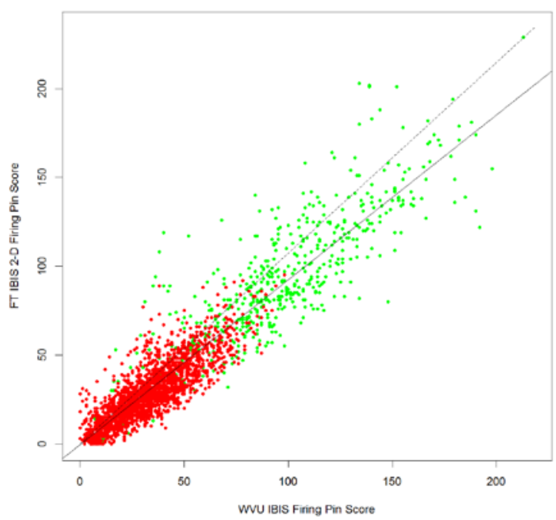

Figure 120: Comparison of breech face scores of the WVU Legacy IBIS and the Forensic Technologies 2D IBIS

Conclusion

A comparison of the 2D Heritage IBIS (upon which this research is based) against that of the new 3D IBIS system (courtesy of Ultra Electronics Forensic Technology Inc., Montreal, Canada) was performed. A selection of twelve 9mm Luger firearms (representing a range of performance characteristics based on IBIS results) was used to produce a set of test cartridge cases. These cartridge cases were run through both systems. The 3D system has a number of advantages most particularly the ability to search the side lit images. Collection of images is more time consuming (±10 minutes) as opposed to the heritage system (±3 minutes). The co-axially illuminated breech face and firing pin images yield similar results in their match scores.

USACIL test set

A test collection of 13 sets of cartridge cases were received from USACIL for testing. Each set contained three known cartridge cases and one questioned cartridge case. All cartridges cases were entered into the IBIS system. All correlation results were processed and the type of each cartridge case was assigned. The types were either Known (Test), Questioned, or Background. Once again the assumption is made that no cartridge cases fired by the same firearms as used in the test set are present in the database. In each instance the Set Number was also included. These data were recorded for both the Sample and Database cartridges. This, for example, allowed for the isolation of Known-versus-Known (test-versus-test) scores for a particular set. The scores for the K_x–K_y pair are the same for the K_y–K_x pair. Thus for the three known cartridge cases there are 3 pairwise comparisons (K_1–K_2, K_1–K_3, K_2–K_3), and for the questioned versus known cartridge cases there are also three comparisons (K_1–Q, K_2–Q, K_3–Q).

Utilization of the AFTE Theory of Identification in the interpretation of IBIS results.

In order to ease of discussion and to provide a clear explanation of the AFTE Theory of Identification (AFTE theory), we will only consider the comparison of cartridge cases as an example.

Through careful examination of cartridge cases firearms examiners and scientists have developed the hypothesis that the markings on the breech face and firing pin of a firearm are transferred to cartridge cases during the discharge of the firearm. After comparison of numerous cartridge cases fired by different firearms, a theory was developed (and continues to be evaluated) for such comparisons. This theory has evolved through an inductive process. A sample, albeit large, of all potential comparisons is used to induce the theory. There may well be two numerically different firearms that mark cartridge cases in the same way. Thus, induction is a probabilistic process of theory development. Examination of more samples, of which the ground truth is known, lends increased support to the theory. Thus, we now have the AFTE theory in its present form. Firearms examiners now use this theory in a deductive fashion (*i.e.* premises and rule (theories, law, etc. guarantee the outcome) in reaching the conclusions as specified in the AFTE theory.

The AFTE theory requires that "sufficient agreement" is required to effect an identification. Significance in comparison is determined "… by the comparative examination of two or more sets of surface contour patterns comprised of individual peaks, ridges and furrows. Specifically, the relative height or depth, width, curvature and spatial relationship of the individual peaks, ridges and furrows within one set of surface contours are defined and compared to the corresponding features in the second set of surface contours."

It may seem atypical, strange even, to use IBIS in a way for which it was not intended. The intention of IBIS was to search through a large number of cartridge cases to identify possible candidates for comparison. The intention of a particular system does not, however, preclude its use for other purposes.

The IBIS system is a tool. A confocal microscope is a tool commonly used to map the surface of some object. The striagraph was a tool to map the surface of a bullet. An atomic force microscope is a tool to map surfaces at extremely high resolutions. A scanning electron microscope (SEM) is a tool to examine surface structures[54]. All of the tools can be used to examine the surface contours of a cartridge case. The question that arises is: "Are they all equal in the way in which they perform?" Clearly, the answer is "no."

The next question is: "How does the tool relate to the theory?" Generally speaking, theories are developed independent of methods. In fact, in certain instances, theories are developed without any measurement (*e.g.* Einstein's theory of general relativity). It would seem, therefore, that the robustness of a theory would be dependent on its ability to be tested by a variety of methods. This is intrinsically a requirement of Daubert, which is based upon the theories of Karl Popper.[55] Subjecting a theory to a "risky" test is an attempt at falsification which is inherent to Popper's theory. Thus, from a scientific stance, it is both prudent and necessary to continue to test any theory. Additionally, a theory which is untestable is deemed to be "junk science."

Using the postulates of the AFTE theory as a basis for the assessment of data generated by the IBIS system, is appropriate in this context.

The use of the IBIS system may, at first gloss, seem to be inherently different from the process of comparison microscopy by an examiner. It is important to differentiate between the process of comparison and that of identification. The images that are viewed by coaxial lighting and side lighting are different, but they are two different representations of the same surface.

[54] At one stage Cambridge Instruments marketed a comparison SEM.
[55] Karl Popper, The Logic of Scientific Discover, 1968.

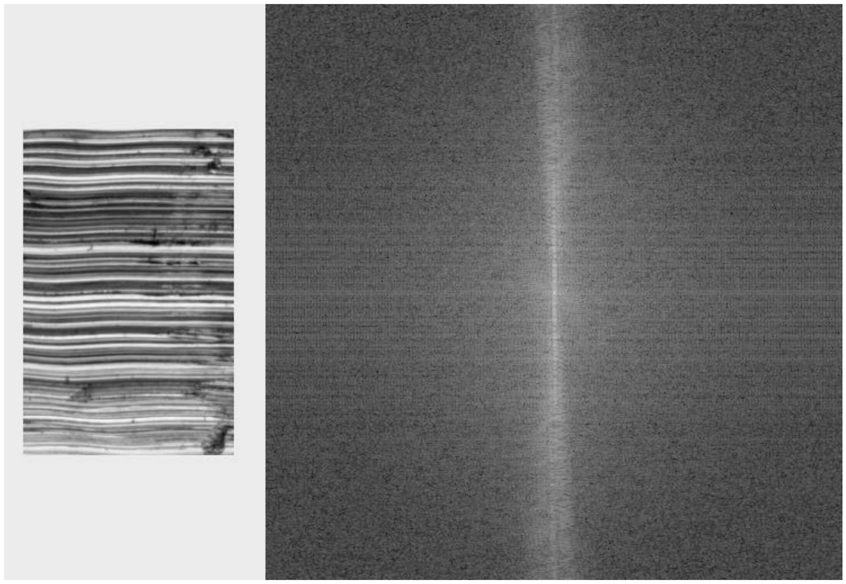

Figure 121: Spatial domain image of a toolmark (left) and frequency domain image of the same toolmark image (right).

Figure 121 provides two representations of the same image. One is in the spatial domain (that which we typically observe) and the other is in the frequency domain. The manner in which images of objects in the spatial domain are compared differs from the way in which frequency domain images of objects are compared. However, these two images represent the same object but in different ways. Therefore, the fact that two methods may be different does not imply that one method is correct and the other incorrect.

In order to effect and identification, the AFTE theory requires “significant agreement.” The agreement is significant when it “… exceeds the best agreement demonstrated between toolmarks known to have been produced by different tools and is consistent with agreement demonstrated by toolmarks known to have been produced by the same tool.” An understanding of this requirement is required to apply to results from the IBIS system.

In this study only the breech face scores and firing pin scores were used. The ejector mark scores could also be added to the interpretation. In order to define the assumptions in this analysis, the following example will be used. A questioned cartridge case was recovered from a crime scene and a suspect was found in possession of a firearm. Three test cartridge cases that were fired in the suspect’s firearm are available for comparison. It is further assumed that if the questioned cartridge case was not fired by the suspect’s firearm then no cartridge cases fired by either of the two firearms are present in the database.[56]

The questioned cartridge case is now submitted to the IBIS database for comparison. The IBIS system will return a list of candidates, which, according to the comparison algorithm, are the

[56] If the questioned cartridge case if was, in fact, fired by the suspect’s firearm then no other cartridge cases fired by that firearm are in the database.

closest matches to the questioned cartridge case. The cartridge case in the candidate list which has the "best" combination of breech face and firing and scores is then the closest candidate. It may also be so that in a particular instance a candidate cartridge case has a well-defined breech face impression and an ill-defined firing pin impression, or vice versa, resulting in low scores for the ill-defined impression.

Since there are no cartridge cases fired by the same firearm as that which fired the questioned cartridge case, all of the cartridge cases in the candidate list are known non-matches (different gun cartridge cases). A plot may be generated, such as that in Figure 124, where the firing pin score is plotted against the breech face score. The maximum product of the firing pin and breech face score of these known non-matches is determined, and this value is assigned as the value of the "best known non-match" as contemplated in the AFTE theory. This may be illustrated as a hyperbola[57] in Figure 124. Thus, if a comparison of the questioned cartridge case and a known cartridge case results in a breech face score and a firing pin score whose product is greater than the best known non-match, then the first part of the section of the AFTE theory under consideration is fulfilled. It must be borne in mind, however, that the utilization of the best known non-match does not imply that a false positive result cannot be made. See Figure 22 to Figure 40 for examples. It must also be remembered that an increase in the nature of the best known non-match will lead to an increase in the false negative rate of the methodology.

If the three known cartridge cases are compared against each other, then these comparisons will form the basis of determining whether or not the questioned cartridge case versus a known cartridge case fulfills the second requirement that the comparison is " … consistent with agreement demonstrated by toolmarks known to have been produced by the same tool." Figure 154 provides a typical example of the situation.

The generally encountered problem is that there is, for the IBIS system at least, significant variability in the distribution of scores that represent the match distribution (or same gun distribution) for the known cartridge cases.[58]

Calculation of likelihood ratios

The Bayesian network given in Figure 122 was used to calculate likelihood ratios for the test and evidence samples. The data given in Table 43 lists the priors (Firing_Pin_Type_Sample and Drag_Mark_Sample) and the evidence (Rank, BF, FP, Rank_BF) for each of the tests. Tests 1, 3, and 7 are similar in that they contain ranks, whilst tests 2 and 4 do not. Each of the tests is conditioned on the firing pin type of the submitted sample (see Table 42). The states of the Firing_Pin_Type_Sample node are *Circular* and *Glock*. The conditioning is necessary to obtain

---

[57] The equation of an hyperbola is given by $x \times y = constant$ – in this case x = breech face score and y = firing pin score.

[58] See Figure 116 as an example of this behavior.

the correct prior odds. It is furthermore necessary that the node Firing_Pin_Type_DB will be conditioned similarly. This assumes that all of the firing pin types in the IBIS are correctly entered. Searching of the IBIS database is conditioned on Firing Pin Type in the system. In other words, when a sample is entered into IBIS, its firing pin type will result in the search being launched against cartridge cases with the same firing pin type. In some instances, the background data were incorrectly classified regarding their firing pin type. Unfortunately, the IBIS does not use the presence or absence of a drag mark as a classifier within its database. Where possible, the presence or absence of a drag mark in the background data has been entered into the test sets. In instances where the state of the drag mark is unknown, the node will be given a state of *Unknown*. The conditioning on drag marks will separate firearms with a blowback action (*e.g.* HiPoint C9) from those with a recoil lock system (*e.g.* Ruger SR9).

Table 42: List of variables used in calculation of likelihood ratios for specified tests

Test Number	Firing_Pin_Type_Sample	Drag_Mark_Sample	BF	FP	Rank	BF_Rank
Test 1	✓		✓	✓	✓	
Test 2	✓		✓	✓		
Test 3	✓	✓	✓	✓	✓	
Test 4	✓	✓	✓	✓		
Test 7	✓	✓	✓	✓	✓	✓

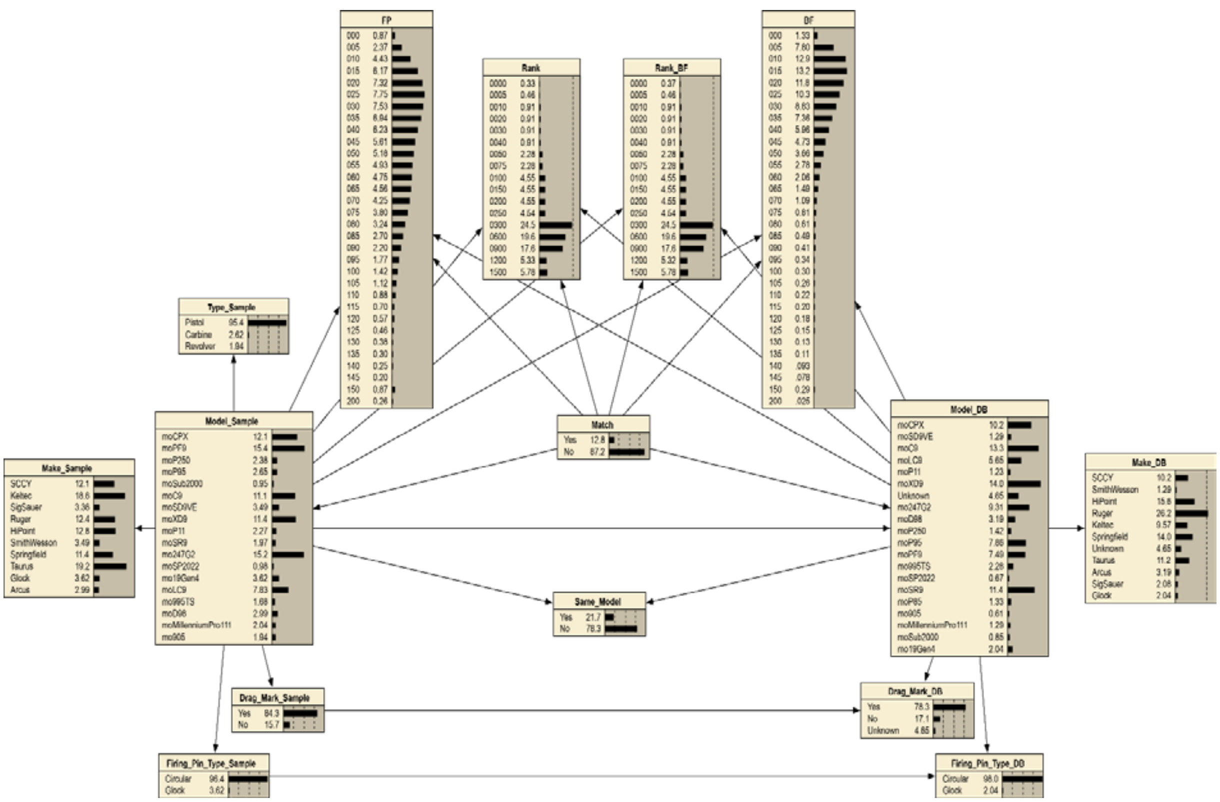

Figure 122: 9mm Bayesian network used for likelihood ratio calculations

Table 43: Classification of Evidence cartridge cases (Priors)

Evidence	Firing Pin Type = Circular Drag Mark = No	Firing Pin Type = Circular Drag Mark = Yes	Firing_Pin_Type = Glock Drag_Mark = Yes
U01	✓		
U02	✓		
U03	✓		
U04			✓
U05	✓		
U06	✓		
U07			✓
U08		✓	
U09		✓	
U10		✓	
U11	✓		
U12		✓	
U13		✓	

U01

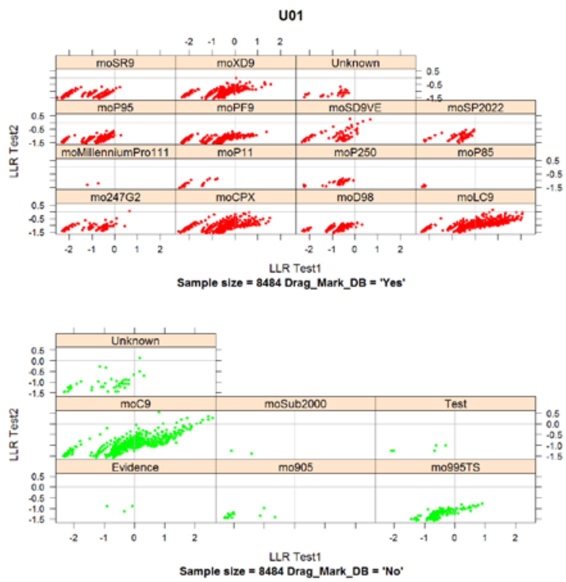

Figure 123: Comparison of the log likelihood ratios (Test1 and Test2) of cartridge case U01

In Figure 123, the logarithm of the likelihood ratios (LLR) for Test 2 are plotted against the LLRs for Test 1. The conditioning factors for these results are given in Table 43, and the classification of the evidence cartridge case is given in Table 44. Both panels of Figure 123 present the data separated by the value of the Model_DB node. The upper panel (and all subsequent similar figures) provides the LLRs with the Drag_Mark_DB node having a state of *Yes*, whilst the lower panel provides the data for the Drag_Mark_DB node having a state of *No*. From Table 44 it is known that the evidence cartridge case, U01, does not feature a drag mark. Thus the plots given in the upper panel represent nonmatching candidates, whilst those in the lower panel represent potentially matching candidates. In this case, the lower panel will also include sections labeled "Test" and "Evidence." The "Test" section provides the LLRs for the test-versus-test samples, whilst the "Evidence" section provides the LLRs for the test-versus-evidence and evidence-versus-test samples. For these results, it must be borne in mind that the results are not conditioned on the Drag_Mark_Sample node.

The LLRs for Test 1 are generally widespread across the zero value. The LLRs for Test 2 are generally below the zero value.

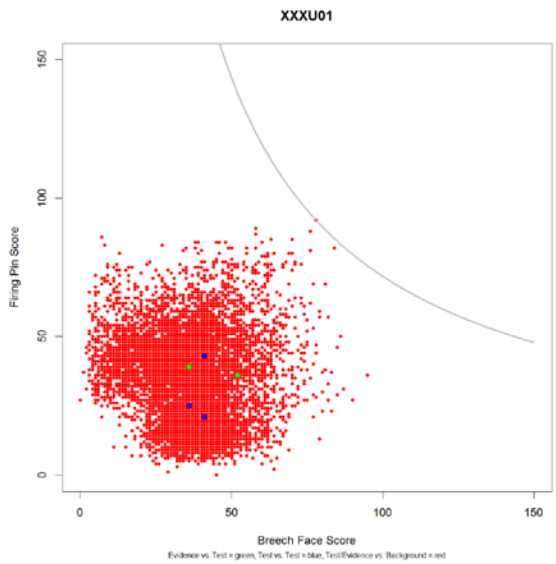

Figure 124: Firing pin versus breech face scores for U01

In Figure 124 (and all subsequent similar plots), the red dots indicate the scores between the evidence and the test samples against the background database. The blue dots indicate the test-versus-test scores, and the green dots indicate the evidence-versus-test scores. The grey curve is the maximum non-match value for the breech face and firing pin scores. In the AFTE theory of identification this would equate to the best-known non-match (BKNM). This plot provides evidence[59] equivalent to that used to calculate the LLR in Test 1.

The three test-versus-test results all lie well within the non-match distribution. Only two of the three evidence-versus-test comparisons (A-1 and A-3) were returned by IBIS. These results also lie well within the non-match distribution.

[59] Evidence (E) in this sense means that as stated in Bayes' theorem (e.g. $\Pr(H_p|E)$)

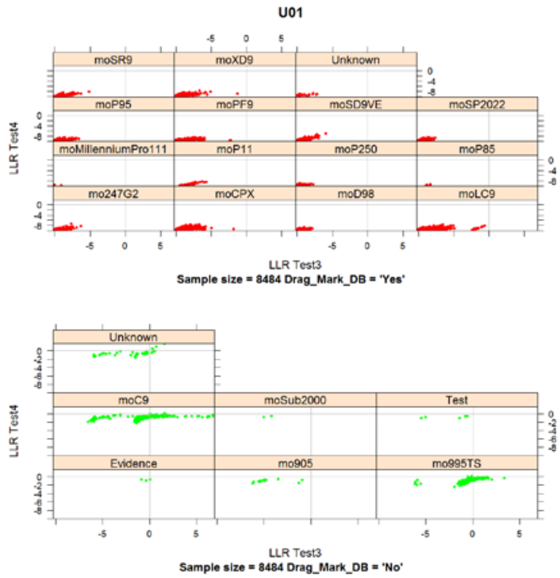

Figure 125: Comparison of the log likelihood ratios (Test3 and Test4) of cartridge case U01

Figure 125 demonstrates the clear difference between those LLRs from the background in the database and the presence/absence of drag marks. Test 3 and Test 4 are different from Test 1 and Test 2 in that the drag mark on the evidence cartridge case is part of the priors for these tests. The LLRs of the background cartridge cases with a drag mark (± 10^{-8}) are significantly lower than those without a drag mark (10^{0}). For both of the tests, the LLRs of the test-versus-test and evidence-versus-test are below zero. The manner in which these LLRs need to be interpreted is as follows. If a firearms examiner deems the evidence and test cartridge cases to have the same characteristics, then the IBIS results can be used to calculate the likelihood ratio based on the database. The IBIS results generate a LLR of approximately 0. This means that evidence cartridge case is as similar to the test cartridge cases as it is to the background cartridge cases based upon the IBIS results. When considering the section for the HiPoint C9 in the lower panel of Figure 125, the LLRs for Test 3 have a wide range (± $10^{-6.5}$ to $10^{6.8}$). For Test 7, the range is significantly smaller (± $10^{-2.1}$ to $10^{0.02}$).

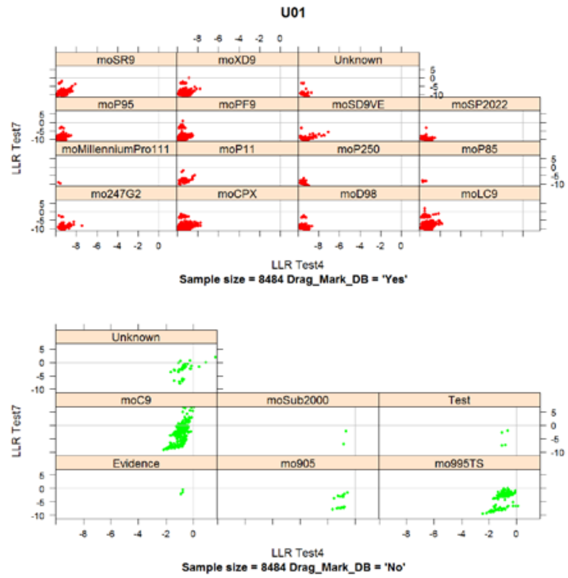

Figure 126: Comparison of the log likelihood ratios (Test4 and Test7) of cartridge case U01

From Figure 126, it can be observed that LLRs both Test 4 and Test 7 for the Test and Evidence sections are less than zero. The LLR are grouped in the same area of the plot. From the HiPoint C9 section there are comparisons which have significantly higher LLRs.

U02

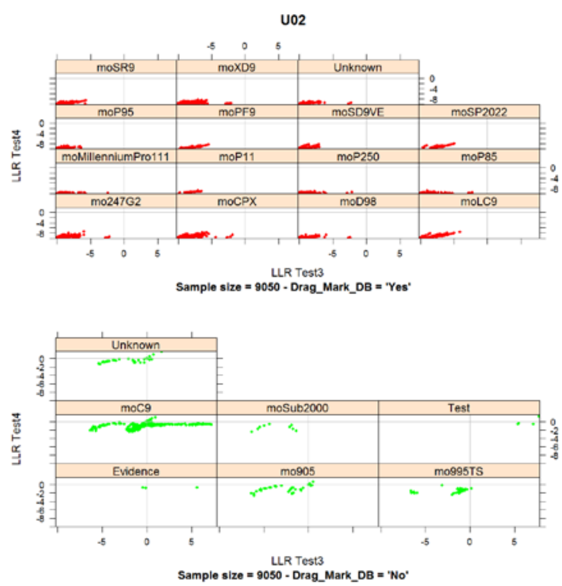

Figure 127: Comparison of the log likelihood ratios (Test3 and Test4) of cartridge case U02

In Figure 127, LLR values for Test 3 for the Evidence of clustered at zero and one at LLR = 5.58. For the Test cartridge cases, the values of the LLR are all greater than five. The higher

LLR for the Evidence is due to a higher firing pin rank (38) while those around zero have firing pin ranks of 410, 645, and 757.

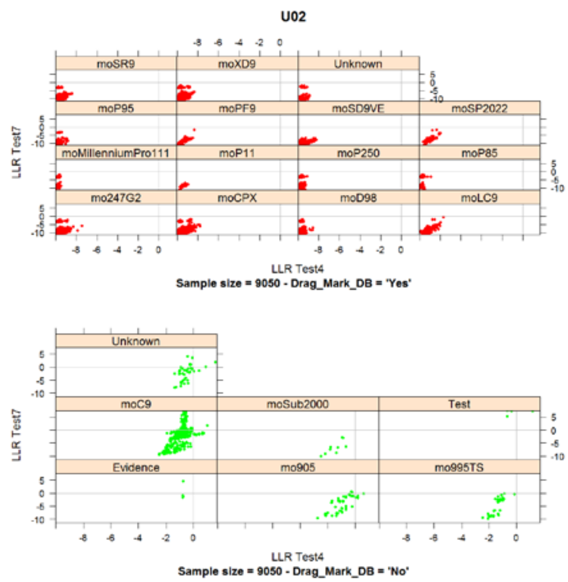

Figure 128: Comparison of the log likelihood ratios (Test4 and Test7) of cartridge case U02

In Figure 128, LLR values for Test 7 for the Evidence are clustered at zero and one at LLR = 4.45. For the Test cartridge cases, the values of the LLR are all greater than five. The higher LLR for the Evidence is due to the firing pin rank, whilst the breech face ranks are only slightly influencing the LLR.

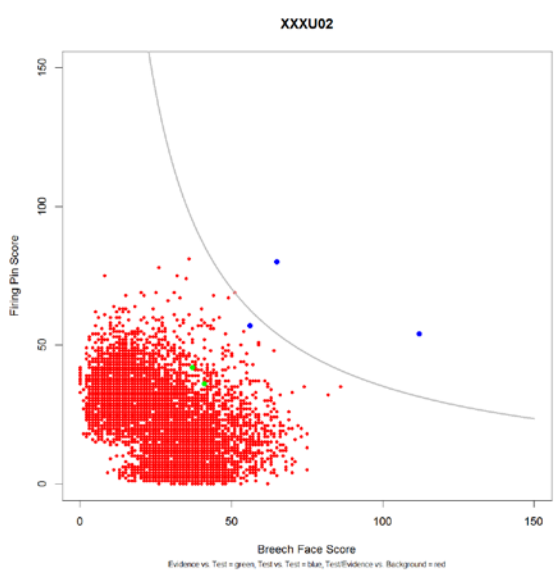

Figure 129: Firing pin versus breech face scores for U02

In Figure 129, two of the tests lie above the BKNM curve. The third, although below the curve, is quite close to it. This clustering of the tests indicates their ability to discriminate against the rest of the database. The results of exhibit B lie within the non-match distribution. Given this evidence the data strongly suggests that exhibit B was not fired from the test firearm. In contrast, the LLRs for Test 1 (see Figure 130) have a maximum of 0.38, whilst the test-versus-test LLRs are the 1.20 to 3.15 range.

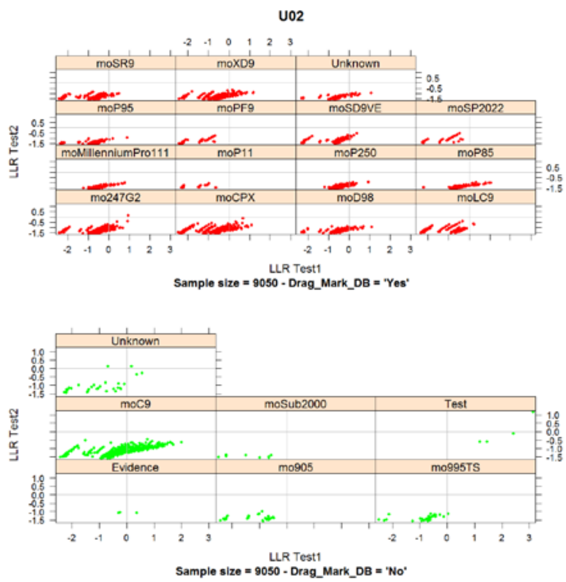

Figure 130: Comparison of the log likelihood ratios (Test1 and Test2) of cartridge case U02

U03

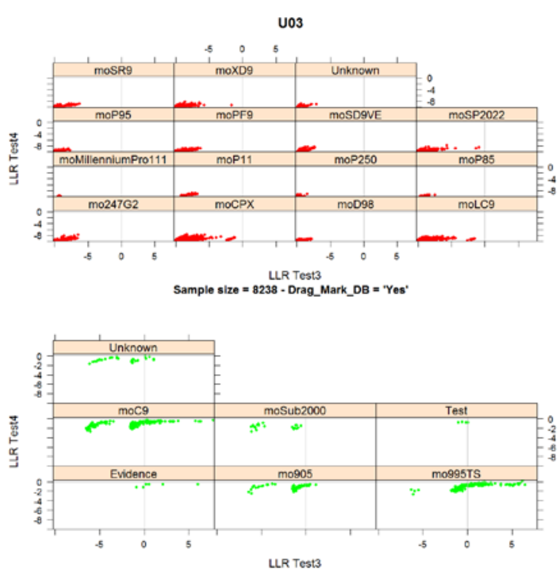

Figure 131: Comparison of the log likelihood ratios (Test3 and Test4) of cartridge case U03

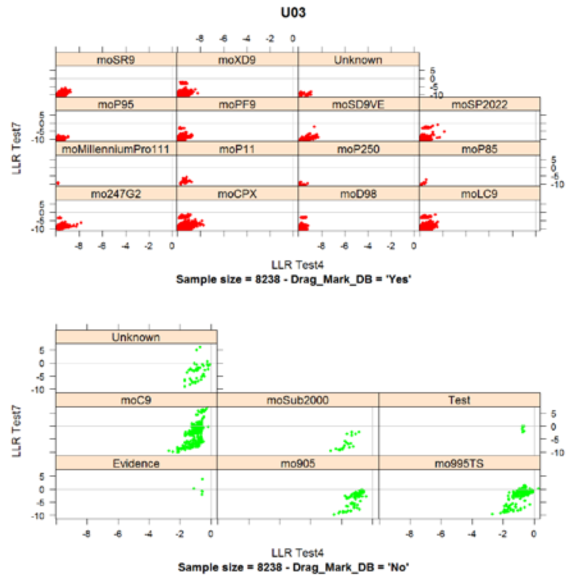

Figure 132: Comparison of the log likelihood ratios (Test4 and Test7) of cartridge case U03

From Figure 131 and Figure 132, it is evident that the LLRs for Test 3 (5.99) and Test 7 (3.71) suggest very strong support for same gun relative to the different gun hypothesis. Test 4 is neutral to very slightly in favor of the different gun hypothesis.

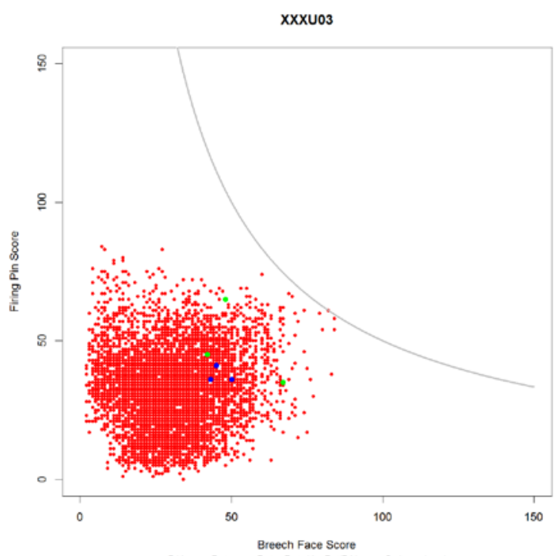

Figure 133: Firing pin versus breech face scores for U03

The three test-versus-test results (blue dots) are all closely clustered well within the non-match distribution. The results of the exhibit C also lie well within the non-match distribution. This indicates that it is not possible for the IBIS system to discriminate these samples from the background of non-matching (different-source) comparisons.

U04

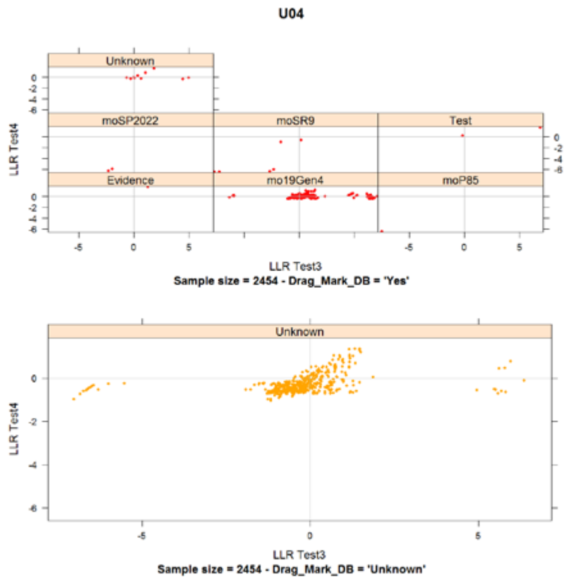

Figure 134: Comparison of the log likelihood ratios (Test3 and Test4) of cartridge case U04

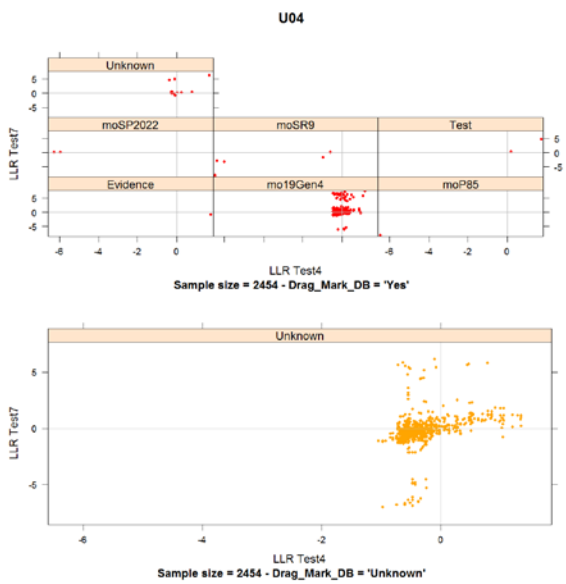

Figure 135: Comparison of the log likelihood ratios (Test4 and Test7) of cartridge case U04

This is the first of the cartridge cases with a Glock type firing pin impression. In this case all candidates in the results will have a Glock type firing pin impression and by design will have a drag mark. For this sample (and sample U07), the lower panel in Figure 134 and Figure 135 represent the unknown firearms with a Glock firing pin impression. In the upper panel, there is a section with a state of the node Model_DB of *Unknown*. These represent cartridge cases which were submitted to IBIS and having Glock type firing pin impressions when, in fact, they were circular. These data will be ignored[60]. In this test, only one test cartridge case provided results against the evidence cartridge case. The resulting LLRs are LLR (Test 3) = 1.25, LLR (Test 4) = 1.76, LLR (Test 7) = -0.77. The first two indicate slight evidence in favor of the same gun hypothesis.

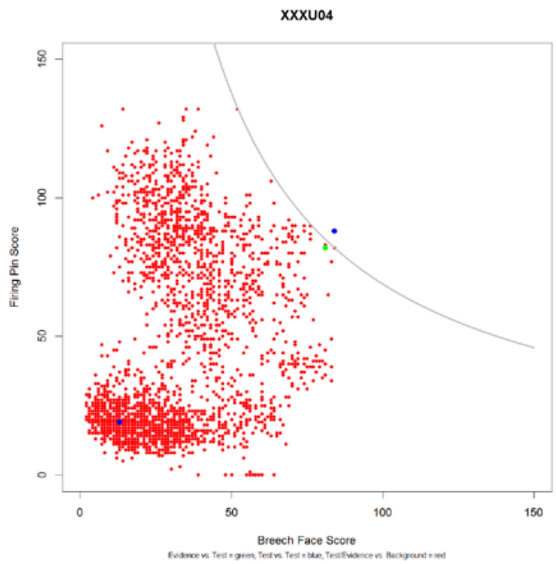

Figure 136: Firing pin versus breech face scores for U04

Figure 136 provides similar support to the LLRs.

[60] These data were later corrected.

U05

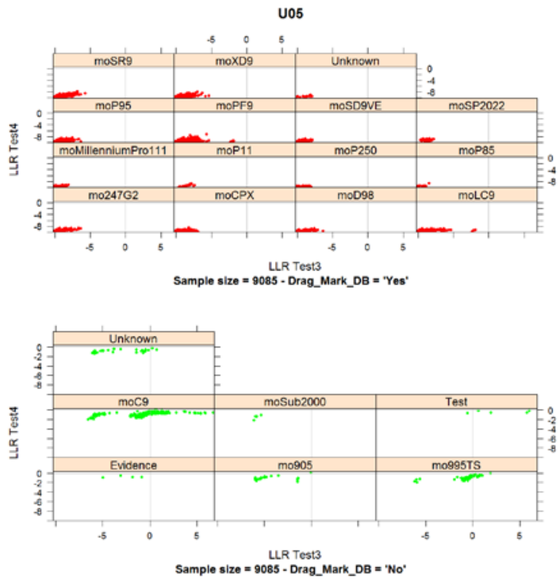

Figure 137: Comparison of the log likelihood ratios (Test3 and Test4) of cartridge case U05

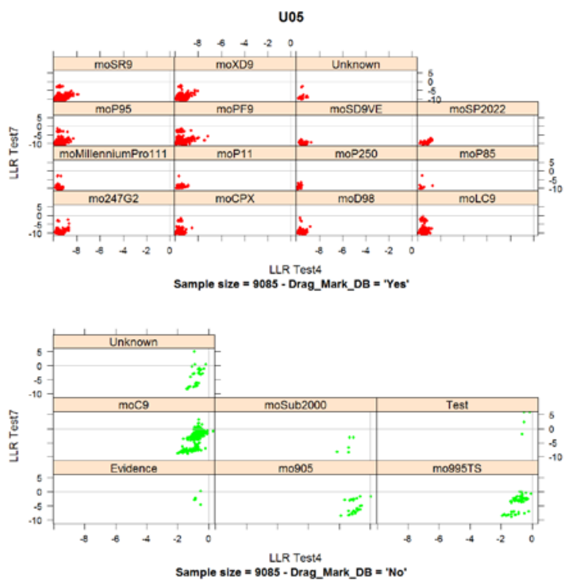

Figure 138: Comparison of the log likelihood ratios (Test4 and Test7) of cartridge case U05

Figure 137 and Figure 138 provide the following ranges for the LLRs: for Test 3 ($-4.96 \leq LLR \leq 0.82$), Test 4 ($-0.92 \leq LLR \leq -0.51$), and Test 7 ($-4.70 \leq LLR \leq 0.23$). All of these data provide strong support for the different gun hypothesis.

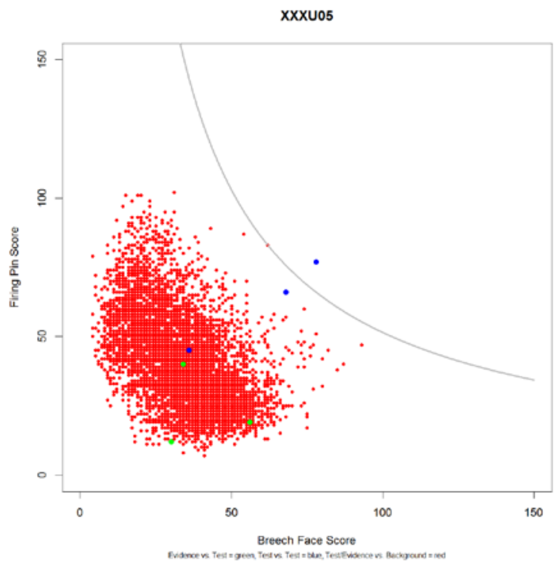

Figure 139: Firing pin versus breech face scores for U05

In Figure 139, one of the tests lies above the BKNM curve. The second lies just below the BKNM curve. The third test lies well within the non-match distribution. Two of the results of exhibit E lie at the lower extreme of the non-match distribution (very low firing pin scores). Given this evidence the data strongly suggests that exhibit B was not fired from the test firearm.

U06

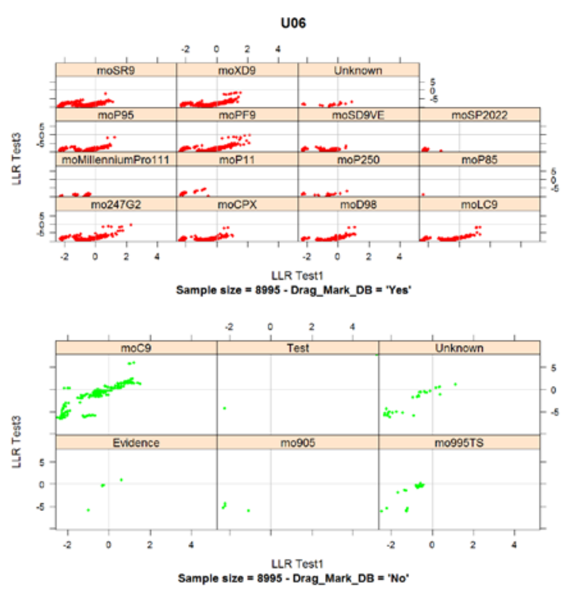

Figure 140: Comparison of the log likelihood ratios (Test1 and Test3) of cartridge case U06

In the Test section of the lower panel of Figure 140 the Test 1 LLR for test cartridge case 16 vs test cartridge case 17 is -2.24 and for test cartridge case 17 versus test cartridge case 18 it is 5.14.

The LLRs for Test 1 for the evidence range from -0.33 to 0.61 (neutral). Test 3 has similar results.

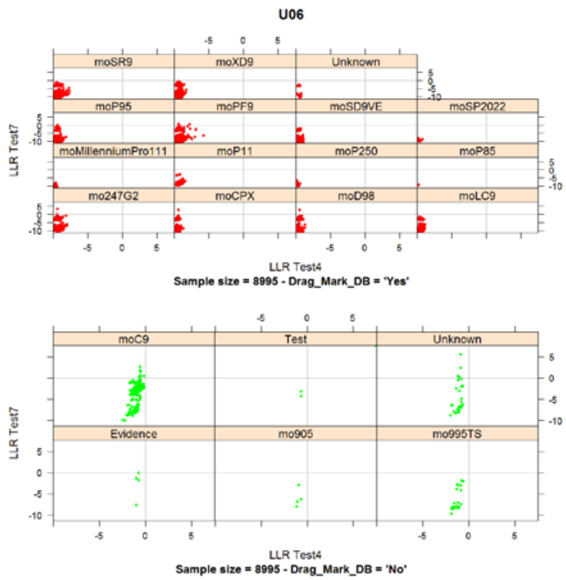

Figure 141: Comparison of the log likelihood ratios (Test4 and Test7) of cartridge case U06

From Figure 141, both Test 4 and Test 7, the LLR provide medium to strong support for the different gun hypothesis.

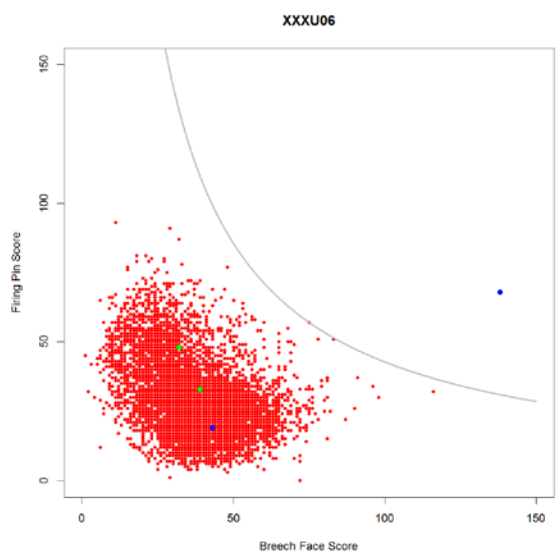

Figure 142: Firing pin versus breech face scores for U06

In Figure 142, one of the tests (T17-T18) lies above the BKNM curve. The second (T16-T17) lies well within the non-match distribution. Only two of the results of exhibit F (F-T17 and F-T18) were returned. Both lie well within the non-match distribution.

U07

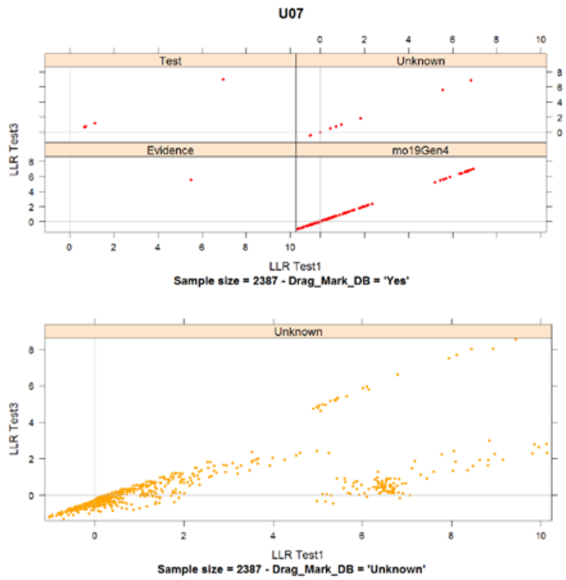

Figure 143: Comparison of the log likelihood ratios (Test1 and Test3) of cartridge case U07

For U07 (Glock type firing pin) all test-versus-test comparisons were returned. Two of the three evidence-versus-test comparisons were returned. Figure 143 and Figure 144 provide LLRs for Test 1, Test 3, and Test 7 between 5.3 and 5.6 (very strong support for the same gun hypothesis). Test 4 returned LLRs of 0.008 and -0.22 (neutral to weak support for the different gun hypothesis). Test 4 is the only test without any rank evidence. Of note, is that a large proportion of all the results have high to very high LLRs.

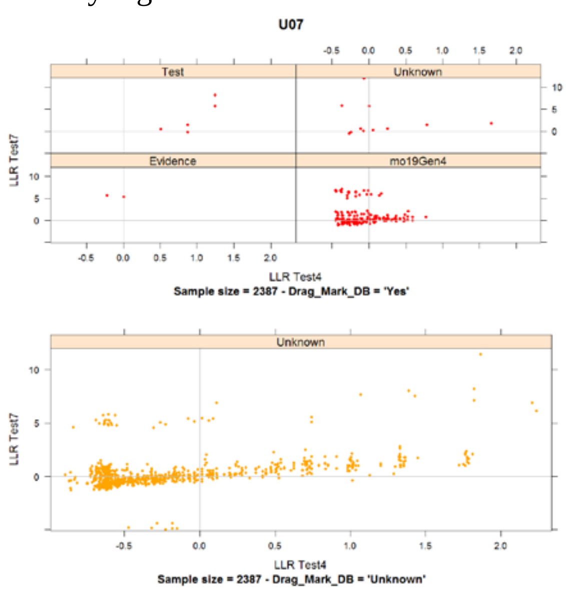

Figure 144: Comparison of the log likelihood ratios (Test4 and Test7) of cartridge case U07

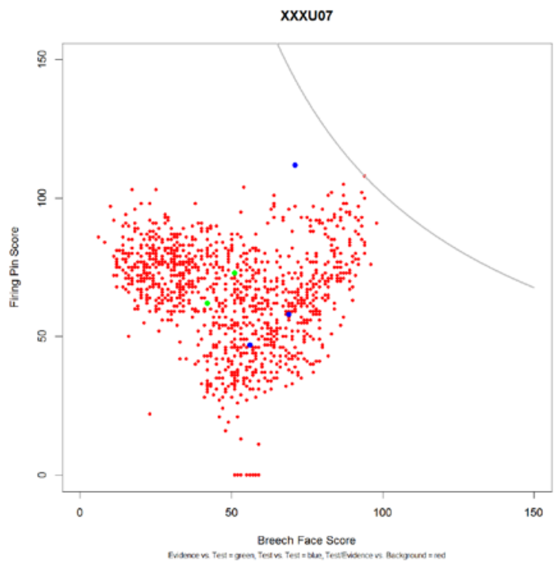

Figure 145: Firing pin versus breech face scores for U07

In Figure 145, none of the tests or the exhibits lie above the BKNM curve, yet some have relatively high LLRs.

U08

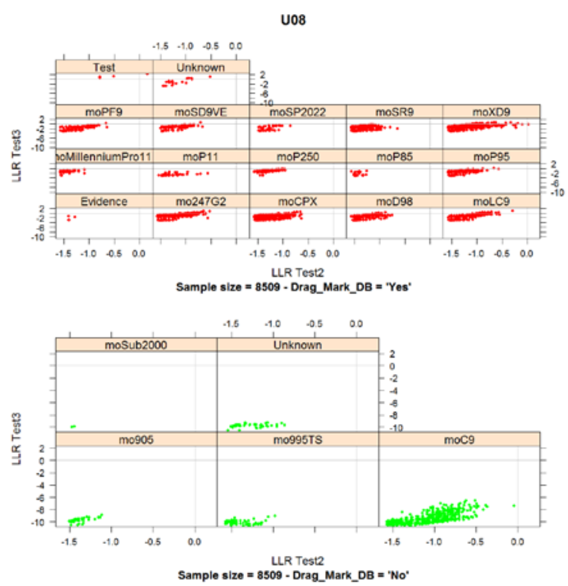

Figure 146: Comparison of the log likelihood ratios (Test2 and Test3) of cartridge case U08

U08 is the first of the recoil action firearms in the test set. It has a circular firing pin impression and a drag mark. In Figure 146, the Test and Evidence sections will be in the upper panel. For Test 1 the LLRs are all below zero except for one against an XD9. For Test 3, the highest LLR for the background is 1.0: for the Evidence the LLRs for Test 1 (-2.28 ≤ LLR ≤ -0.57), Test 3 (-3.02 ≤ LLR ≤ -1.23), Test 4 (-1.32 ≤ LLR ≤ -1.19), and Test 7 (-3.37 ≤ LLR ≤ -1.49).

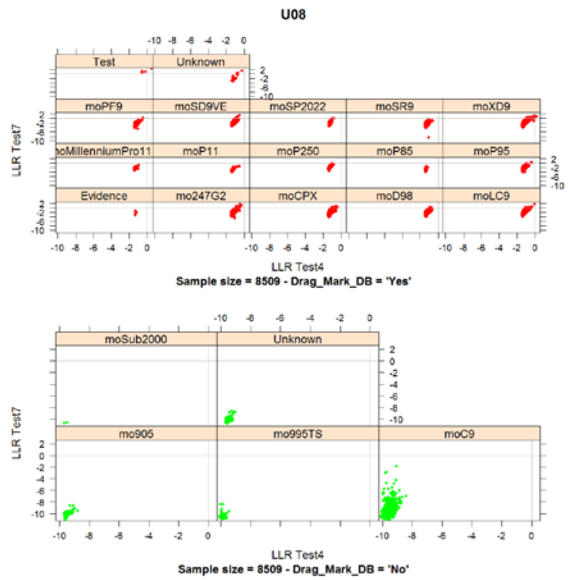

Figure 147: Comparison of the log likelihood ratios (Test4 and Test7) of cartridge case U08

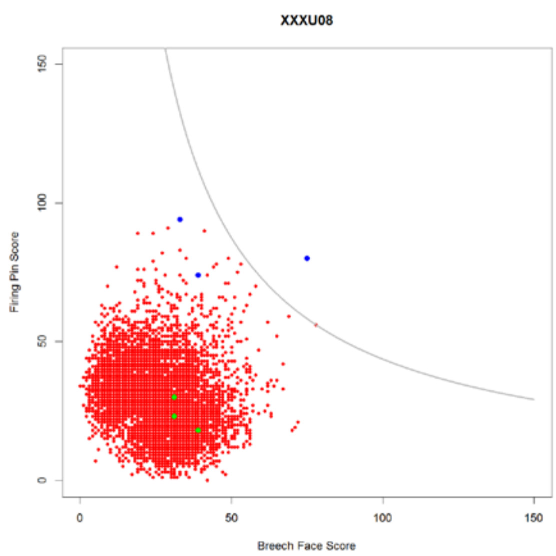

Figure 148: Firing pin versus breech face scores for U08

In Figure 148, one of the tests (T22-T24) lies above the BKNM curve. The second and third, although below the curve, is quite close to it. One test (T23-T24) has the highest firing pin score of all. This clustering of the tests indicates their ability to discriminate against the rest of the database. The results of exhibit G lie well within the non-match distribution. Given this evidence the data strongly suggests that exhibit G was not fired from the test firearm. This is supported by the LLRs.

U09

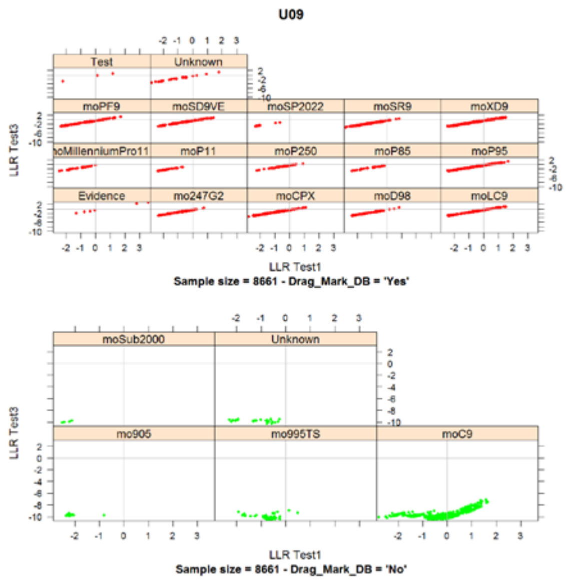

Figure 149: Comparison of the log likelihood ratios (Test1 and Test3) of cartridge case U09

From Figure 149, there appears to be a linear relationship between the LLRs for Test 1 and Test 3. Test 3 scales over a greater range (added sample drag mark): for the Evidence the LLRs for Test 1 ($-1.30 \leq LLR \leq 3.61$), Test 3 ($-2.02 \leq LLR \leq 2.96$), Test 4 ($-1.00 \leq LLR \leq 0.94$), and Test 7 ($-2.55 \leq LLR \leq 3.48$). This provides an indication of the variability of the test cartridge cases as well as strong support for the same gun hypothesis.

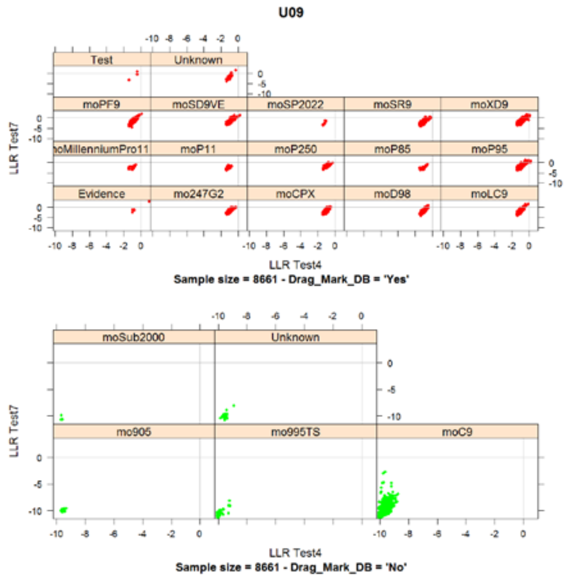

Figure 150: Comparison of the log likelihood ratios (Test4 and Test7) of cartridge case U09

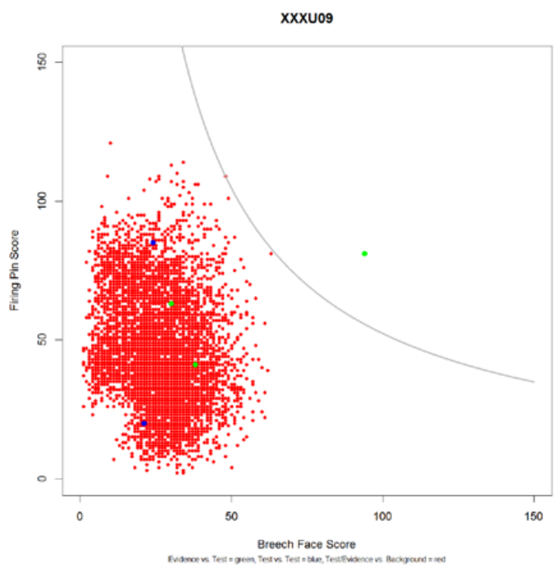

Figure 151: Firing pin versus breech face scores for U09

In Figure 151, the two returned test results have similar breech face scores (21 vs. 24), but quite different firing pin scores (20 vs. 85). The tests scores (T25-T26 and T25-T27) lie well below the BKNM curve. The T26-T27 pair did not even return a score indicating its weak performance. This indicates that these two pairs of cartridge cases do not represent the firearm very well. The single exhibit score is well above the BKNM curve is (I-27). Given this interpretation, the exhibit was fired by the same firearm but the firearm exhibits high variability in its marking.

U10

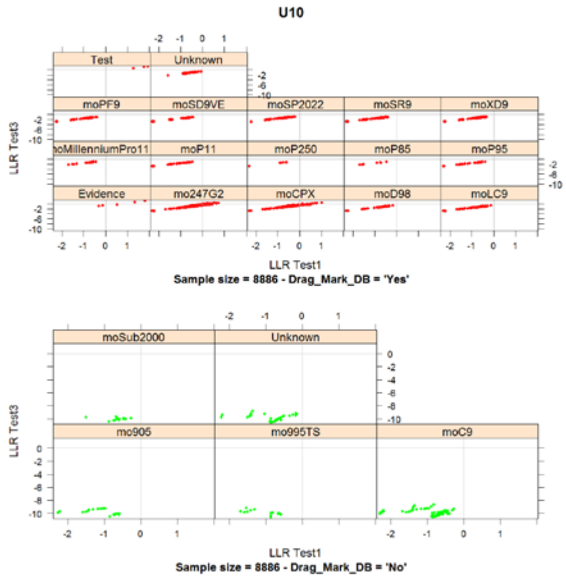

Figure 152: Comparison of the log likelihood ratios (Test1 and Test3) of cartridge case U10

In Figure 152 and Figure 153, the LLRS are tightly group with the Evidence and Tests having the highest relative scores of all for Test 1, Test 3, Test 4, and Test 7. Only Test 4 resulted in negative LLRs. The LLRs for the other tests provide medium to strong support for the same gun hypothesis.

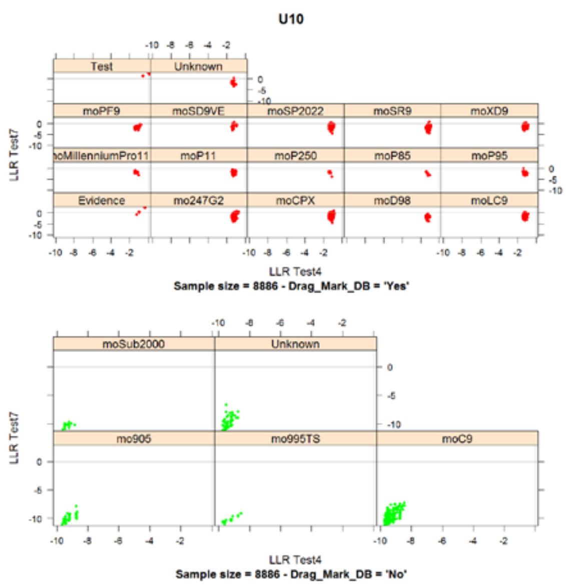

Figure 153: Comparison of the log likelihood ratios (Test4 and Test7) of cartridge case U10

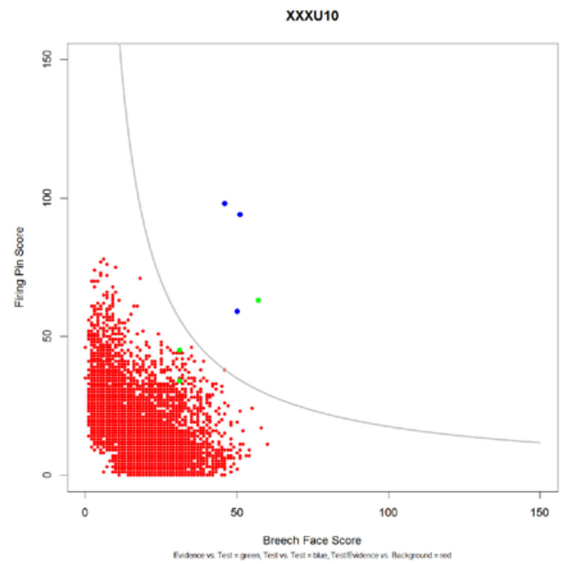

Figure 154: Firing pin versus breech face scores for U10

In Figure 154, all of the tests lie well above the BKNM curve. This indicates that all three of the cartridge cases are well representative of the firearm. One of the exhibit results test (J-T30) is also above the BKNM curve. The other two exhibit results, although below the BKNM curve, are quite close to. Given this evidence the data strongly suggests that exhibit G was fired from the test firearm.

U11

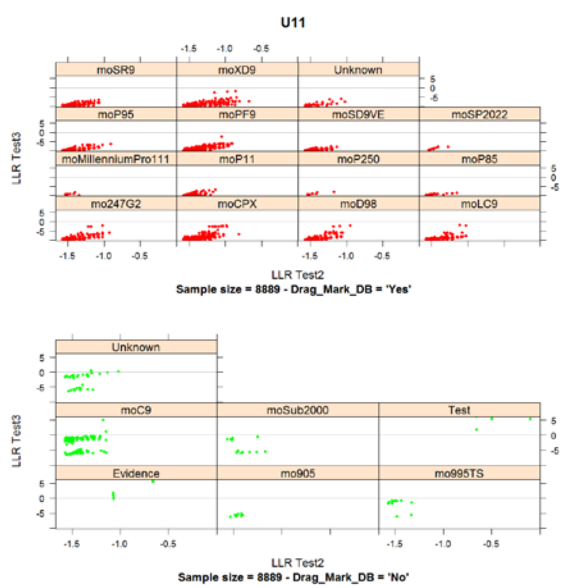

Figure 155: Comparison of the log likelihood ratios (Test2 and Test3) of cartridge case U11

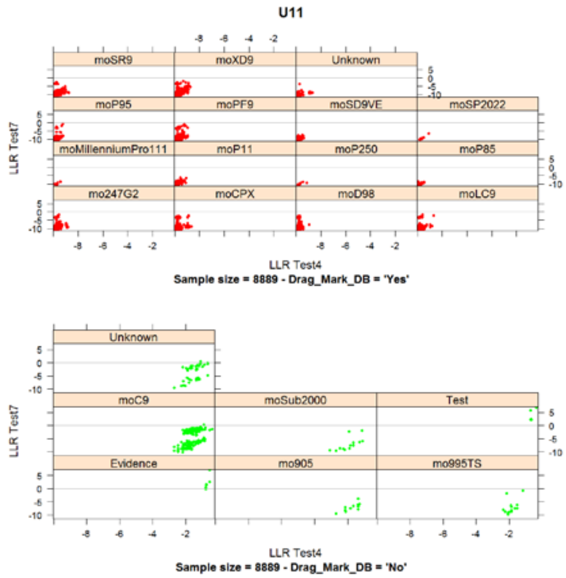

Figure 156: Comparison of the log likelihood ratios (Test4 and Test7) of cartridge case U11

In Figure 155 and Figure 156, Test 3 (-0.13 ≤ LLR ≤ 6.14) and Test 7 (-0.15 ≤ LLR ≤ 7.12) provide weak to extremely strong evidence in support of the same gun hypothesis, whilst Test 2 and Test 4 provide weak evidence in support of the different gun hypothesis.

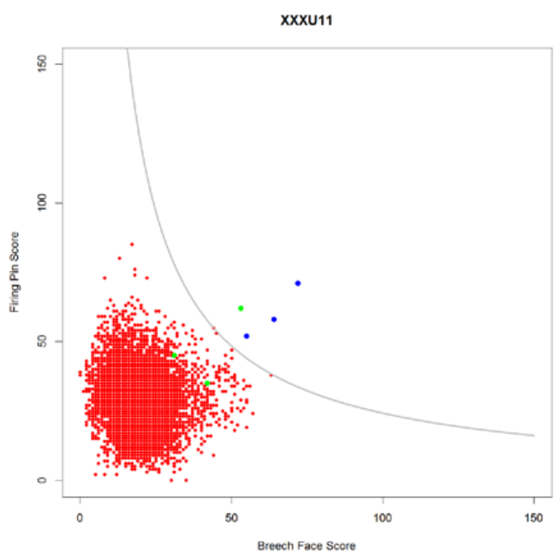

Figure 157: Firing pin versus breech face scores for U11

In Figure 157, all of the tests lie well above the BKNM curve. This indicates that all three of the cartridge cases are well representative of the firearm. One of the exhibit results test (K-T32) is also above the BKNM curve. The other two exhibit results, although below the BKNM curve, are also quite close. Given this evidence the data strongly suggests that exhibit K was fired from the test firearm.

U12

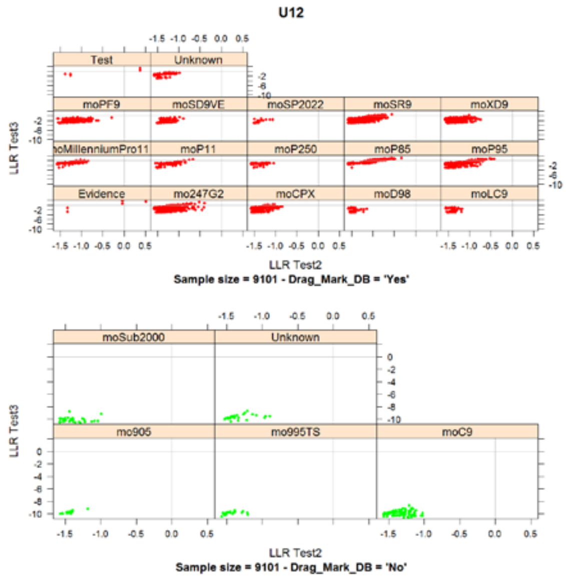

Figure 158: Comparison of the log likelihood ratios (Test2 and Test3) of cartridge case U12

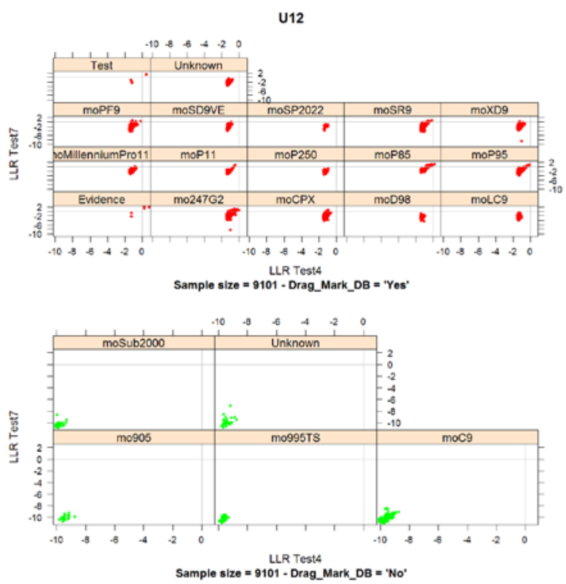

Figure 159: Comparison of the log likelihood ratios (Test4 and Test7) of cartridge case U12

In Figure 158 and Figure 159, the tests indicate the following ranges for the LLRs: Test 2 ($-1.33 \leq LLR \leq 0.51$), Test 3 ($-2.75 \leq LLR \leq 2.04$), Test 4 ($-1.24 \leq LLR \leq 0.80$), and Test 7 ($-2.36 \leq LLR \leq 2.48$) provide a range of support to both hypotheses. It appears that, generally, the absence of ranks provides weaker support for the same gun hypothesis than when the ranks are included in the LLR calculations.

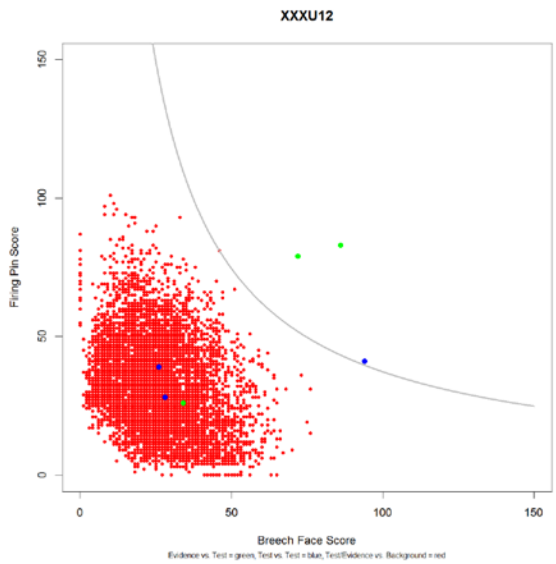

Figure 160: Firing pin versus breech face scores for U12

In Figure 160, one of the tests (T35-T36) lies just above the BKNM curve. The other two are well within the non-match distribution. This indicates that these cartridge cases do not represent the firearm as weel as the first pair. Two of the exhibit results (L-T34 and L-T35) are also well above the BKNM curve. The other exhibit result is also well within the non-match distribution. Given this evidence the data strongly suggests that exhibit L was fired from the test firearm.

U13

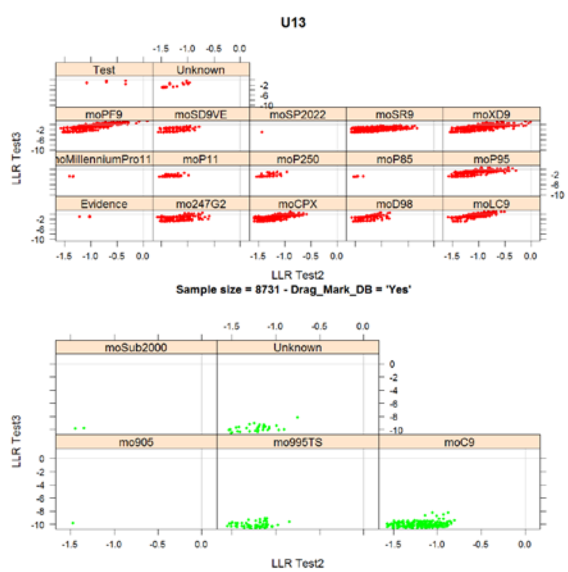

Figure 161: Comparison of the log likelihood ratios (Test2 and Test3) of cartridge case U13

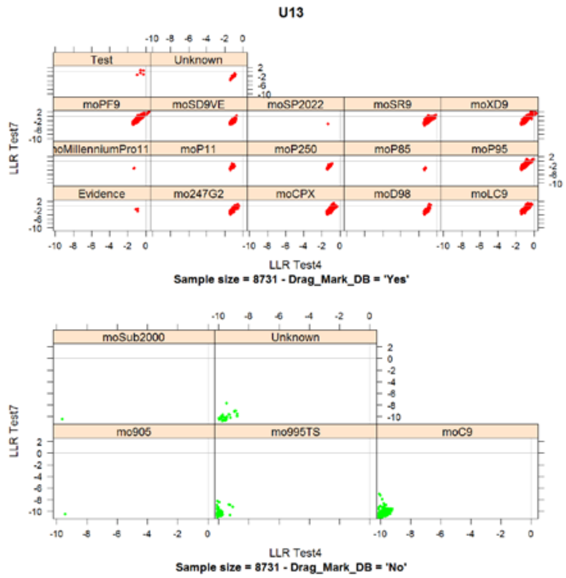

Figure 162: Comparison of the log likelihood ratios (Test4 and Test7) of cartridge case U13

From Figure 161 and Figure 162, it can be seen that Test 4 and Test 7 indicate relatively neutral LLRs for the Test samples. In all cases the Tests indicate strong to weak support for the different gun hypothesis.

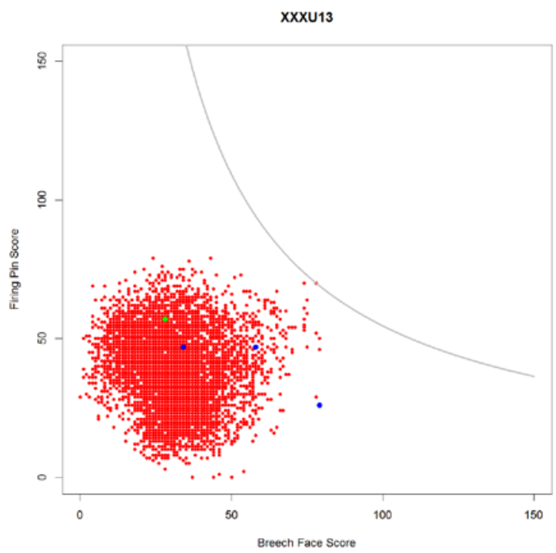

Figure 163: Firing pin versus breech face scores for U13

Three test-versus-test results were returned (Figure 163), of which two lie well within the non-match distribution. The third pair (T37-T38) has the highest breech face score. Only one of the three evidence-versus-test comparisons (M-T39) was returned by IBIS. This result also lies well within the non-match distribution.

USACIL test set revisited

Table 44: USACIL Test Set - firearms information

Sample Set	Known Firearm Make/Model
U01	Sig Sauer P228
U02	Sig Sauer P226
U03	Sig Sauer P226
U04	Glock 19
U05	Ruger P89DC
U06	Ruger P89DC
U07	Glock 19
U08	Smith &Wesson SW9VE
U09	Smith &Wesson SW9VE
U10	Taurus PT 24/7 PRO
U11	Taurus PT 24/7 PRO
U12	Taurus PT 709
U13	Springfield Armory XDM-9

After receipt of the information given in Table 44, a reassessment of the data provided resulted in the adaption of the Bayesian network to differentiate between the presence of a drag mark on the prime of a cartridge case and the type of action of the firearm. Generally, two main types of pistol actions are encountered within the data set. Blowback action is a type of design in which there is no locking of the bolt. The breech is held closed only by the weight and inertia of the bolt, with some slight assistance from the recoil spring, until the bullet leaves the muzzle[61]. In a recoil action (locked breech) pistol, the barrel and slide are securely locked together at the moment of firing. They travel backward together until the barrel unlocks, forced down by a link or inclined plane, and continues rearward under its own momentum[62]. A HiPoint C9 pistol has a blowback action, whilst a Ruger SR9 has a recoil action. Drag mark are generally only found on cartridges fired by a recoil action pistol. Some recoil action pistols seldom generate a drag mark on their cartridge cases e.g. SigSauer P250.

[61] Nonte, G.C., Firearms encyclopedia, Harper & Row, 1973, p. 29.
[62] Nonte, G.C., Firearms encyclopedia, Harper & Row, 1973, p. 208.

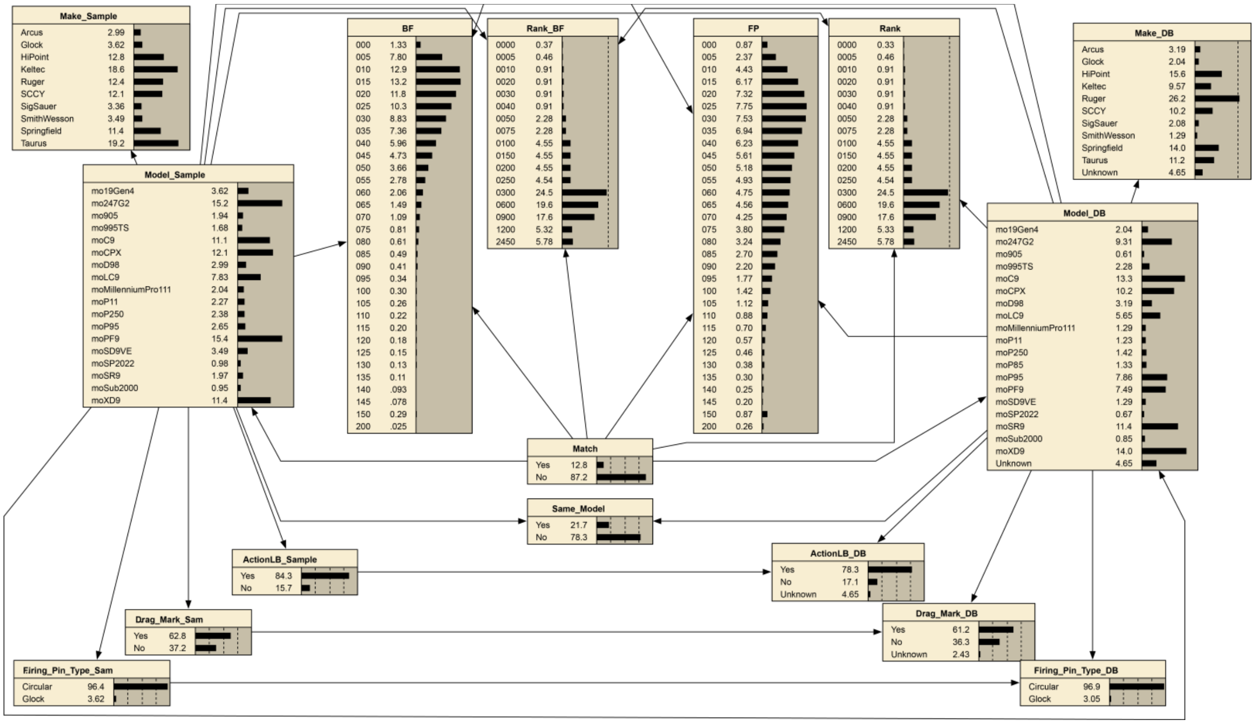

Figure 164: Bayesian network updated to accommodate action type and presence of a drag mark

The effect of this change is illustrated in Figure 165. A SCCY CPX II is selected as the model of the firearm. This pistol has a recoil action and thus has a locked breech. The Yes state of the node ActionLB_Sample becomes 100%. When the Match node is instantiated to Yes, the ActionLB_DB updates to Yes =100%. A match can only be between the same SCCY CPX II pistol, which are a locked breech action. For the nodes Drag_Mark_Sample = Yes (42.4%) and Drag_Mark_Sample = No (57.6%) indicating that the presence of drag marks on these samples is not well replicated. The inference to be made is that if a fired cartridge case was found from the SCCY CPX II pistol there is a 42.4% probability that it will have a drag mark.

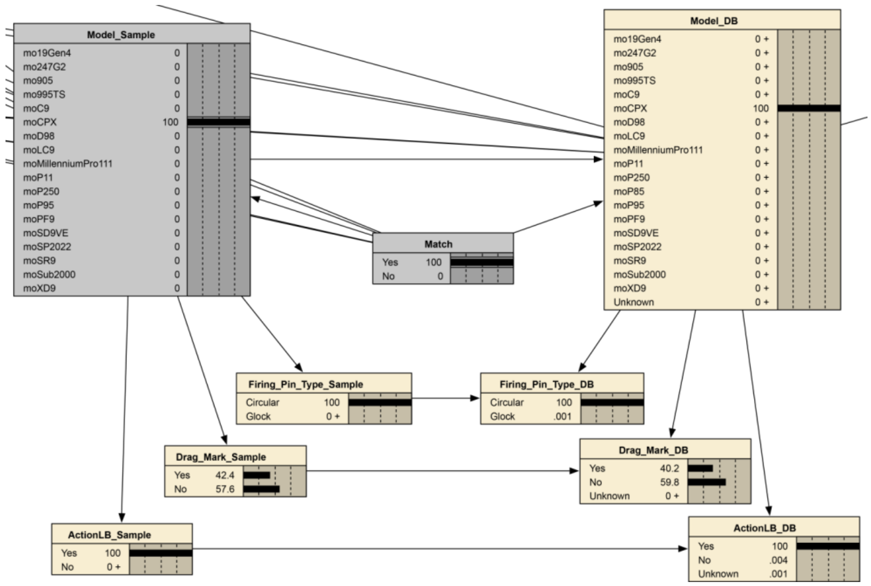

Figure 165: Update illustrating the operation of the nodes Drag_Mark and Action_LB

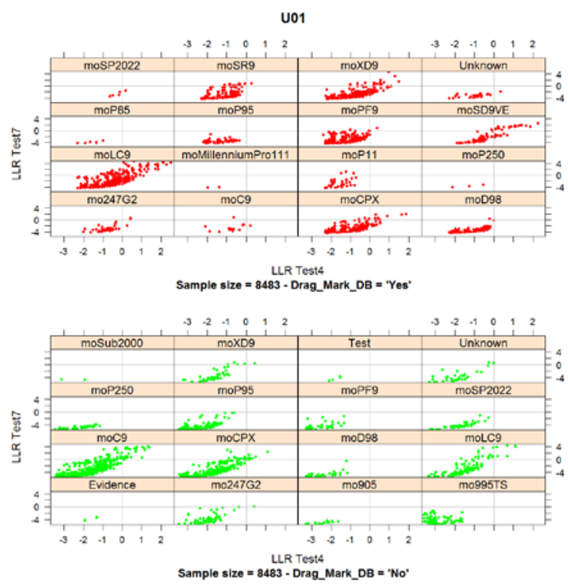

Figure 166: Comparison of the log likelihood ratios (Test4 and Test7) of cartridge case U01 using the updated Bayesian network by Drag_Mark_DB (Sig Sauer P228)

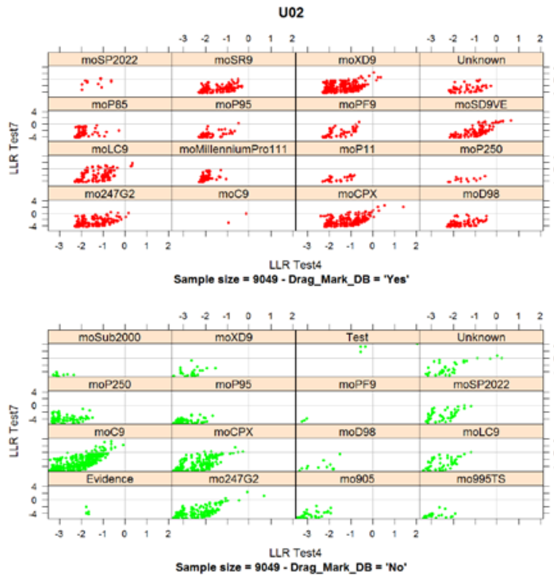

Figure 167: Comparison of the log likelihood ratios (Test4 and Test7) of cartridge case U02 using the updated Bayesian network by Drag_Mark_DB (Sig Sauer P226)

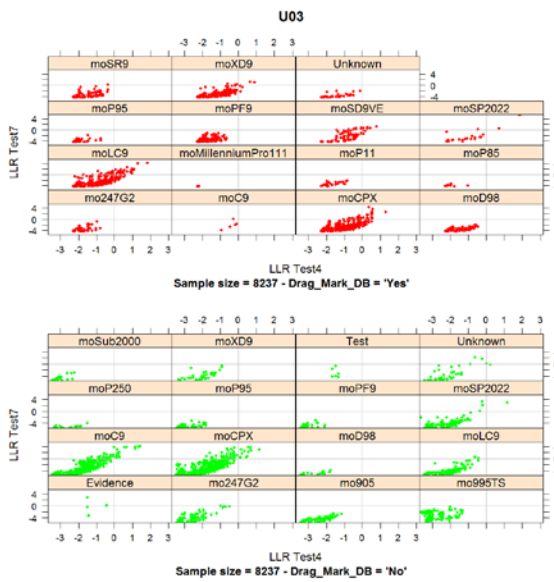

Figure 168: Comparison of the log likelihood ratios (Test4 and Test7) of cartridge case U03 using the updated Bayesian network by Drag_Mark_DB (Sig Sauer P226)

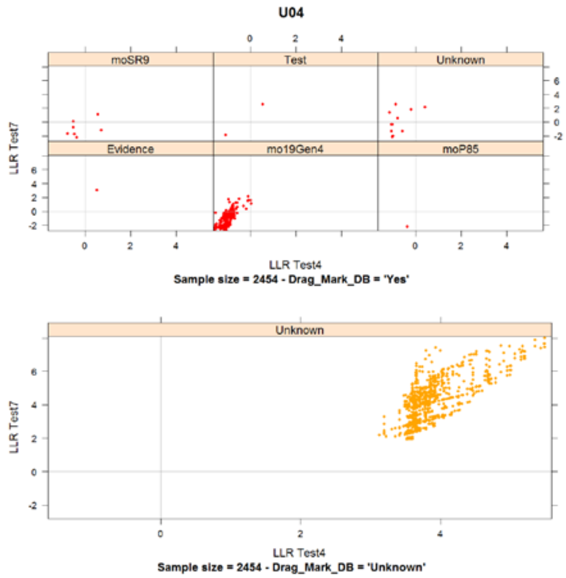

Figure 169: Comparison of the log likelihood ratios (Test4 and Test7) of cartridge case U04 using the updated Bayesian network by Drag_Mark_DB (Glock 19)

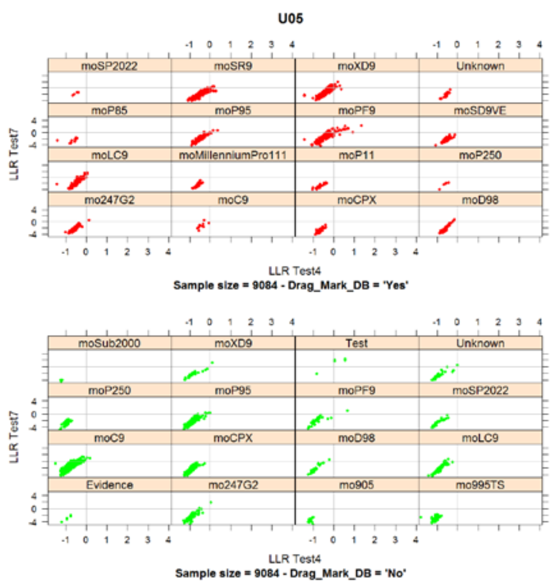

Figure 170: Comparison of the log likelihood ratios (Test4 and Test7) of cartridge case U05 using the updated Bayesian network by Drag_Mark_DB (Ruger P89DC)

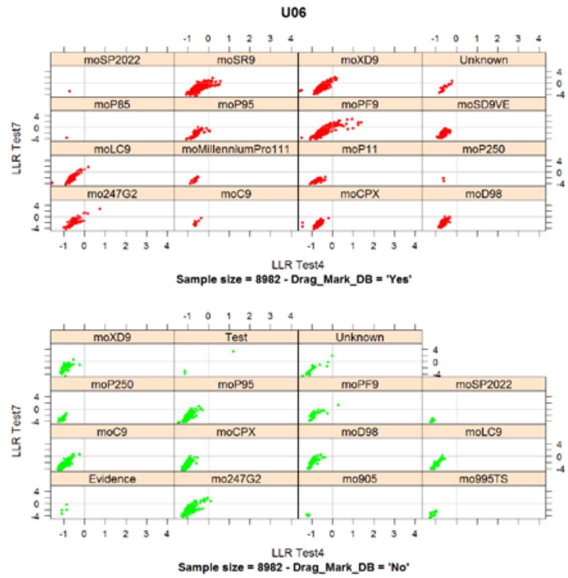

Figure 171: Comparison of the log likelihood ratios (Test4 and Test7) of cartridge case U06 using the updated Bayesian network by Drag_Mark_DB (Ruger P89DC)

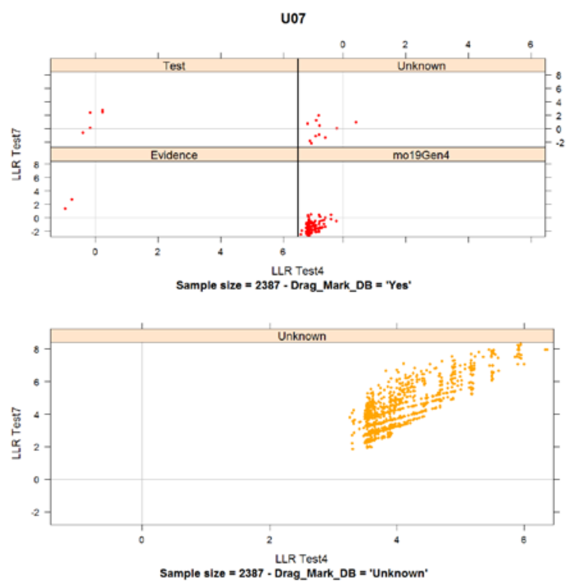

Figure 172: Comparison of the log likelihood ratios (Test4 and Test7) of cartridge case U07 using the updated Bayesian network by Drag_Mark_DB (Glock 19)

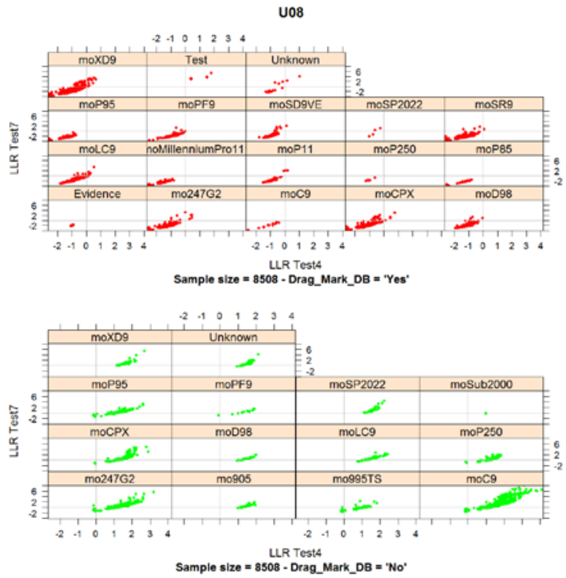

Figure 173: Comparison of the log likelihood ratios (Test4 and Test7) of cartridge case U08 using the updated Bayesian network by Drag_Mark_DB (Smith &Wesson SW9VE)

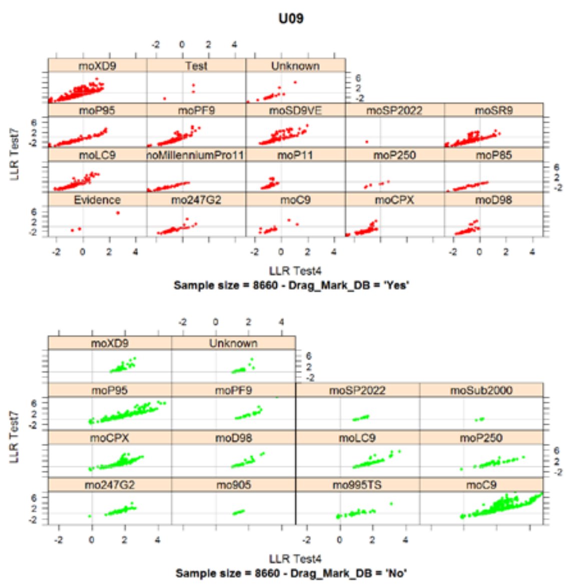

Figure 174: Comparison of the log likelihood ratios (Test4 and Test7) of cartridge case U09 using the updated Bayesian network by Drag_Mark_DB (Smith &Wesson SW9VE)

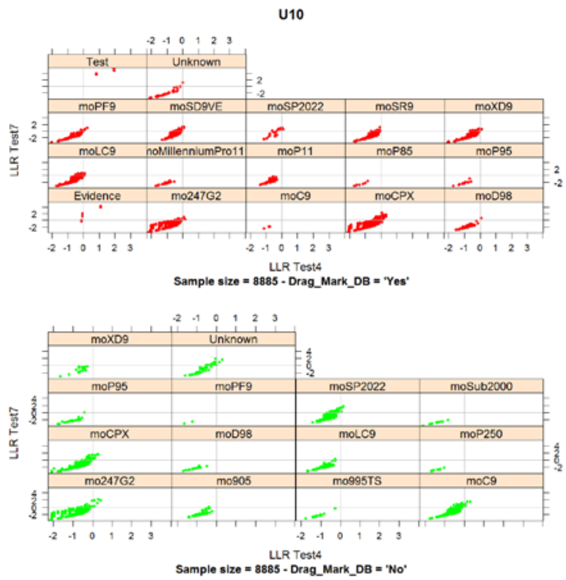

Figure 175: Comparison of the log likelihood ratios (Test4 and Test7) of cartridge case U10 using the updated Bayesian network by Drag_Mark_DB (Taurus PT 24/7 PRO)

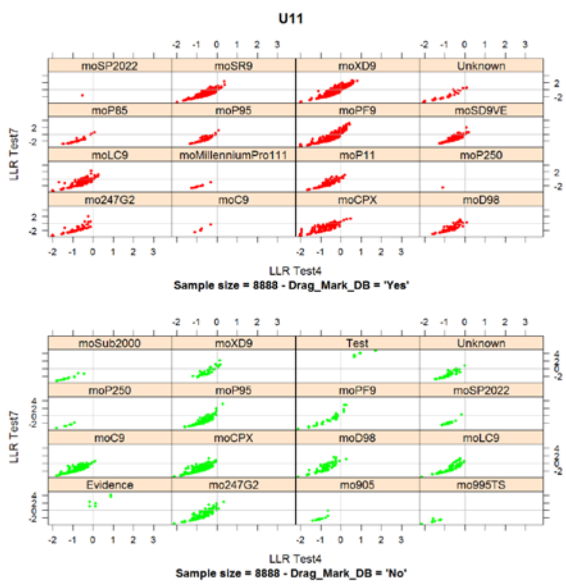

Figure 176: Comparison of the log likelihood ratios (Test4 and Test7) of cartridge case U11 using the updated Bayesian network by Drag_Mark_DB (Taurus PT 24/7 PRO)

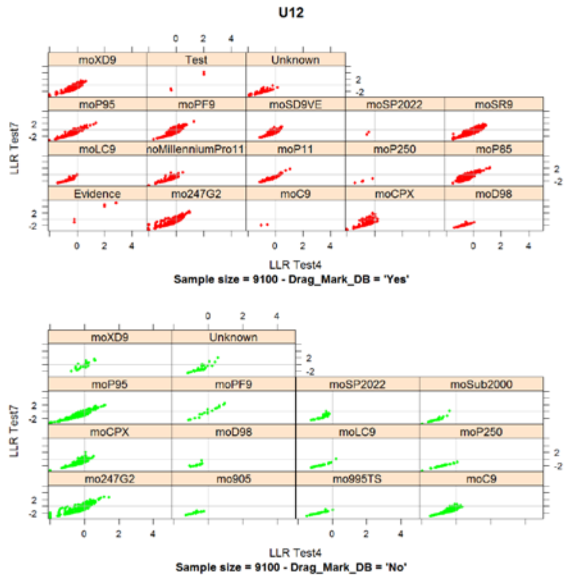

Figure 177: Comparison of the log likelihood ratios (Test4 and Test7) of cartridge case U12 using the updated Bayesian network by Drag_Mark_DB (Taurus PT 709)

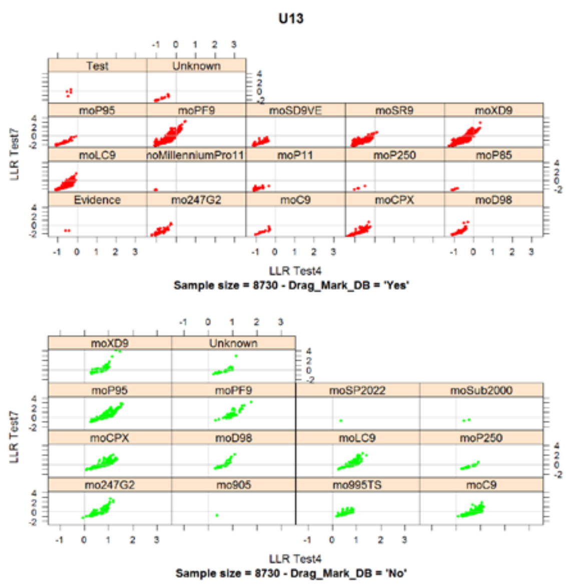

Figure 178: Comparison of the log likelihood ratios (Test4 and Test7) of cartridge case U13 using the updated Bayesian network by Drag_Mark_DB (Springfield Armory XDM-9)

Conclusion

A test set was received from USACIL and analyzed using the developed Bayesian Networks. An assessment of the data is provided. After discussions with the program manager the make and model of each firearm was provided. The test cartridge cases from the SigSauer pistols did not leave drag marks, but were of the recoil action type. The Bayesian network was updated to reflect this information. The drag mark node was split into two, viz. Drag_Mark_Sample(DB) and ActionLB_Sample(DB). Thus from a prior odds perspective the sample would (or not) have a drag mark, but could be from a recoil action pistol.

Baldwin test set

In a study conducted by Baldwin et al. 25 Ruger SR9 pistols were conditioned by firing 200 cartridges in each pistol. Thereafter 800 cartridges were fired through each pistol and collected. Sets of one "questioned" cartridge case and three "known" cartridge cases were set up by the Baldwin group and sent out to firearms examiners for further analysis. Twenty sets were selected by the Defense Forensic Science Center (DFSC), Office of the Chief Scientist and submitted for analysis. The purpose of this test was to study a set of cartridge cases which had previously been examined by a group of practicing firearms examiners in an effort to assess false-positive and false-negative rates in cartridge case comparisons[63]. Each test set contained three known cartridge cases and one questioned cartridge case. The ground truth and the results of examination of the cartridges cases were withheld until completion of the study. The procedure for analysis was the same as described under the section "USACIL Test Set Revisited."

Table 45: Maximum LLRs for all Baldwin data

		Max LLR						
Sample	**Number of Records**	**LLR Test 1**	**LLR Test 2**	**LLR Test 3**	**LLR Test 4**	**LLR Test 7**	**Max LLR**	**Verbal Scale Value**
Set 01	6	2.00	-0.09	1.37	0.08	1.61	2.00	Evidence strongly supports Hp
Set 02	1	-0.69	-1.28	-1.41	-1.24	-1.65	-0.69	Evidence weakly supports Hd
Set 03	6	3.58	2.05	2.74	2.18	2.86	3.58	Evidence very strongly supports Hp
Set 04	2	-0.97	-1.38	-1.69	-1.34	-1.88	-0.97	Evidence weakly supports Hd
Set 05	1	0.33	-0.95	-0.41	-0.93	0.50	0.50	Evidence weakly supports Hp
Set 06	4	2.01	-0.45	1.46	-0.37	2.29	2.29	Evidence strongly supports Hp
Set 07	4	1.42	-0.46	0.72	-0.41	1.66	1.66	Evidence supports Hp
Set 08	4	2.15	0.15	1.57	0.38	2.22	2.22	Evidence strongly supports Hp
Set 09	0							
Set 10	3	0.66	0.10	-0.07	0.13	1.29	1.29	Evidence supports Hp
Set 11	0							
Set 12	9	2.37	0.40	1.89	0.70	1.96	2.37	Evidence strongly supports Hp
Set 13	5	2.20	0.29	1.63	0.51	1.53	2.20	Evidence strongly supports Hp
Set 14	4	-0.15	-1.28	-0.87	-1.24	-0.52	-0.15	Evidence weakly supports Hd
Set 15	6	1.33	-0.84	0.60	-0.79	1.61	1.61	Evidence supports Hp
Set 16	6	3.67	1.64	2.90	1.62	3.07	3.67	Evidence very strongly supports Hp
Set 17	0							
Set 18	3	0.74	-0.69	0.02	-0.62	1.26	1.26	Evidence supports Hp
Set 19	6	2.73	0.78	2.23	1.09	2.69	2.73	Evidence strongly supports Hp
Set 20	2	1.90	-0.21	1.27	-0.03	1.15	1.90	Evidence supports Hp

[63] David P. Baldwin, Stanley J. Bajic, Max Morris, and Daniel Zamzow. *A Study of False-Positive and False-Negative Error Rates in Cartridge Case Comparisons*, Ames Laboratory, USDOE, Technical Report # IS-5207, April 7, 2014 funded through the Office of the Chief Scientist, Defense Forensic Science Center.

Table 45 provides the results of the determination of the log likelihood ratios (LLR) for the evidence versus test samples in each of the sets. The number of records returned indicates the test/evidence comparisons that were returned by IBIS. For Set 09, Set 11, and Set 17 no records were retuned. In these data, all of the records from the Ruger SR9 study previously entered into IBIS were removed from the candidate lists and the firing pin and breech face ranks were recalculated without those data. In the plots the Model DB of unknown contains all comparison data between sets. For this analysis no prior information regarding the test firearms has been considered (i.e. the make and model of the gun is unknown).

Set 01

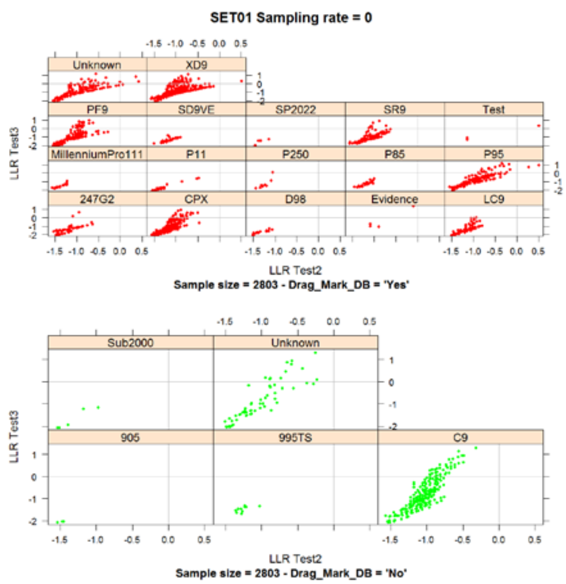

Figure 179: Comparison of the log likelihood ratios (Test2 and Test3) of evidence cartridge case from Set 01

Figure 179 and Figure 180 provide the LLRs of Test 2, Test 3, Test 4, and Test 7 for each test cartridge case in the test set. These data are separated by the presence of a drag mark and the model of the firearm in the database. These figures are the same for the rest of the samples. The Test block indicates test-versus-test cartridge cases and the Evidence block indicates LLRs for evidence against test cartridge cases. These figures are the same for the rest of the samples.

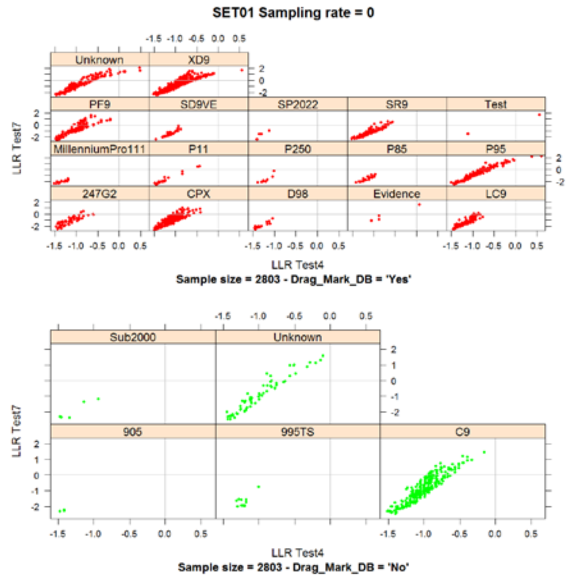

Figure 180: Comparison of the log likelihood ratios (Test4 and Test7) of evidence cartridge case from Set 01

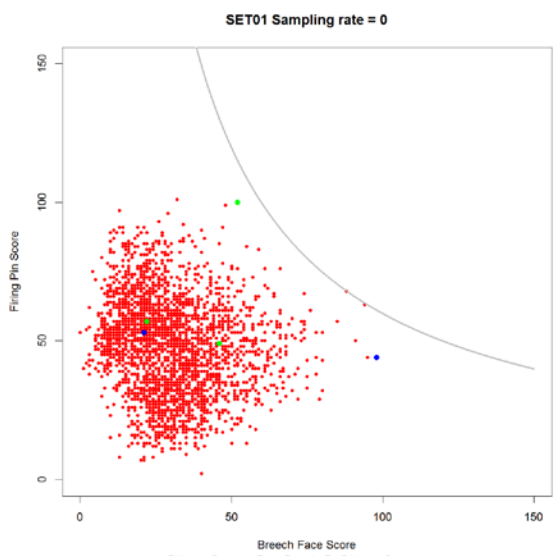

Figure 181: Firing pin versus breech face scores for Set 01

Figure 181 provides an example of the data for set 01. These are the raw data scores from the IBIS system and are generally used to assess a preliminary match status of the evidence. The green dots represent the evidence-versus-test scores, while the blue dots represent test-versus-test scores (an indication of the reproducibility of marketing within the file. The solid curve represents the best-known non-match (BKNM) curve. This curve is developed from the background data that were returned by both the test and evidence cartridge cases. Two of the evidence-versus-test scores were well within the background data, but one is just below the best-known non-match. The data may support an outcome of a match.

Set 02

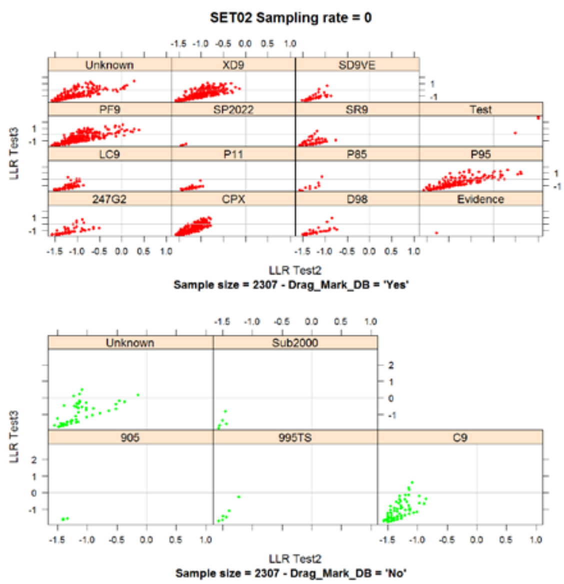

Figure 182: Comparison of the log likelihood ratios (Test2 and Test3) of evidence cartridge case from Set 02

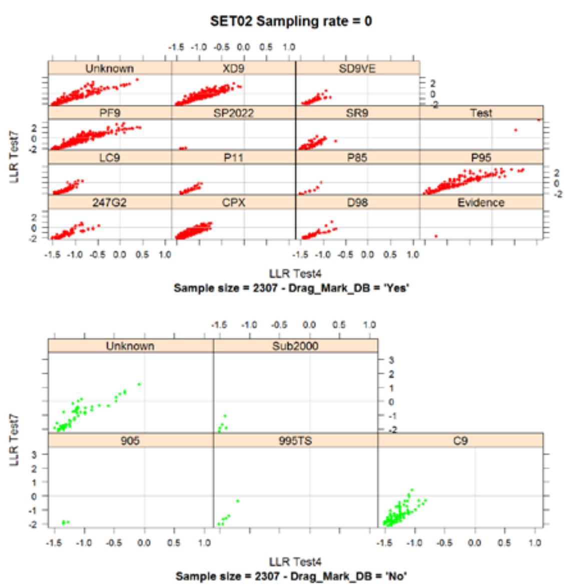

Figure 183: Comparison of the log likelihood ratios (Test4 and Test7) of evidence cartridge case from Set 02

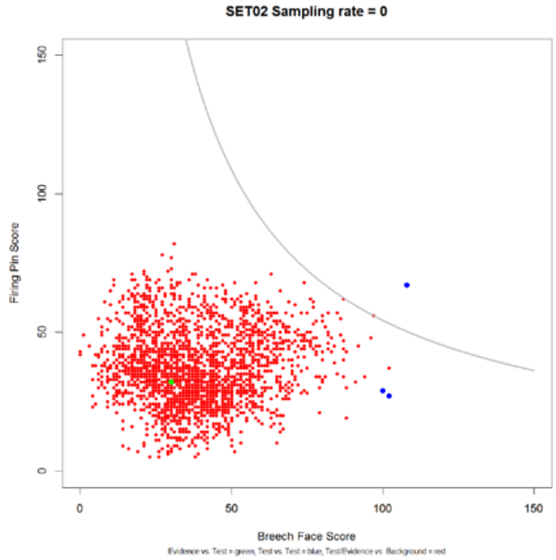

Figure 184: Firing pin versus breech face scores for Set 02

Figure 184 represents the results for Set 03. The reproducibility of the test samples is quite high with the breech face scores. The scores for the test samples are about the position for the highest non-match breech face score. One test comparison is well beyond the best-known non-match score line. There is only one evidence-versus-test score available that lies well within the background data. Using this information this result is most likely a non-match.

Set 03

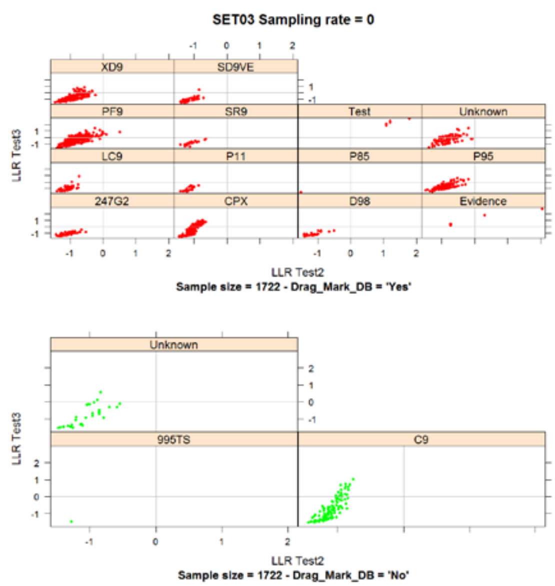

Figure 185: Comparison of the log likelihood ratios (Test2 and Test3) of evidence cartridge case from Set 03

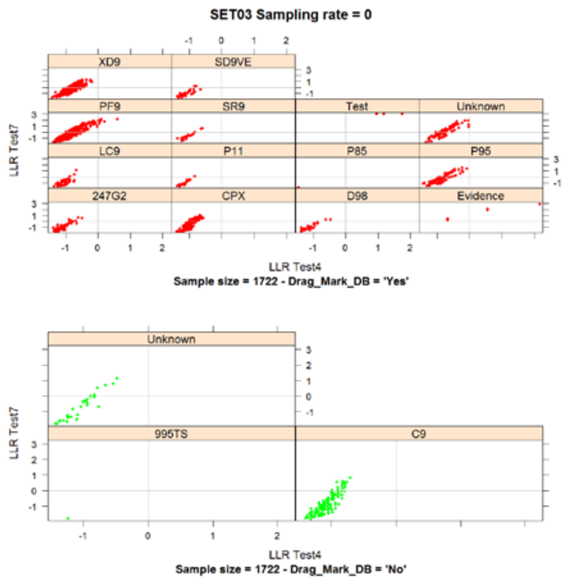

Figure 186: Comparison of the log likelihood ratios (Test4 and Test7) of evidence cartridge case from Set 03

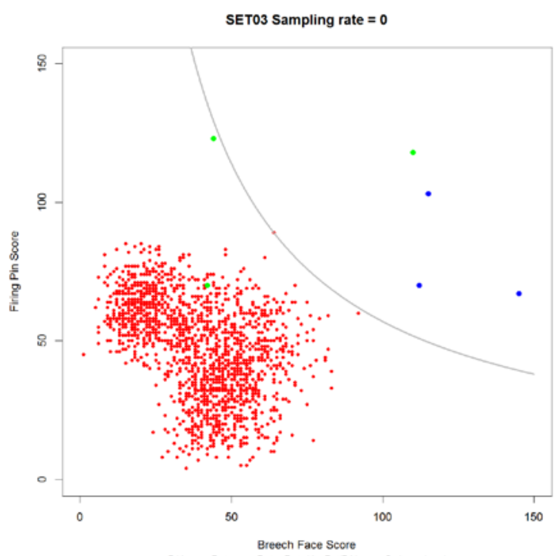

Figure 187: Firing pin versus breech face scores for Set 03

Figure 187 represents the data for Set 03. The test-versus-test scores are all well above the best known non-match. One evidence-versus-test score is at the outer periphery of the background data. A second is just below the BKNM line, but has a very high firing pin score which supports the same gun hypothesis. The final evidence-versus-test score is well above the best-known non-match curve, exceeding that of the test-versus-test scores. These data strongly support a match.

Set 04

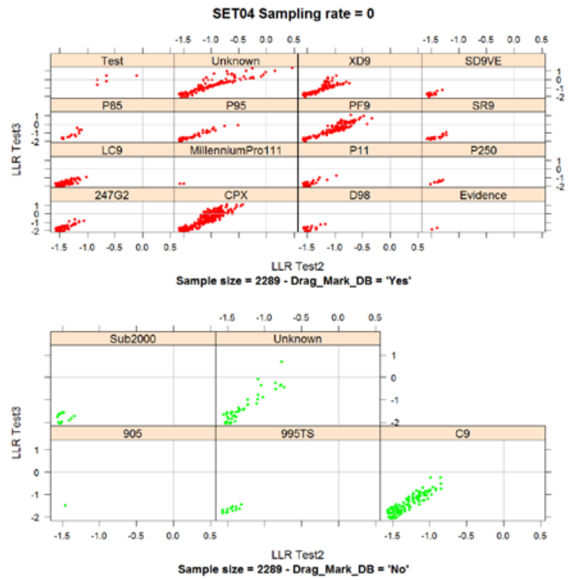

Figure 188: Comparison of the log likelihood ratios (Test2 and Test3) of evidence cartridge case from Set 04

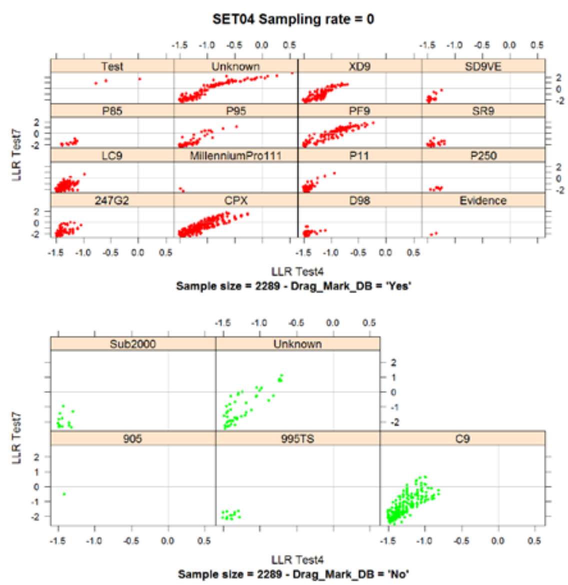

Figure 189: Comparison of the log likelihood ratios (Test4 and Test7) of evidence cartridge case from Set 04

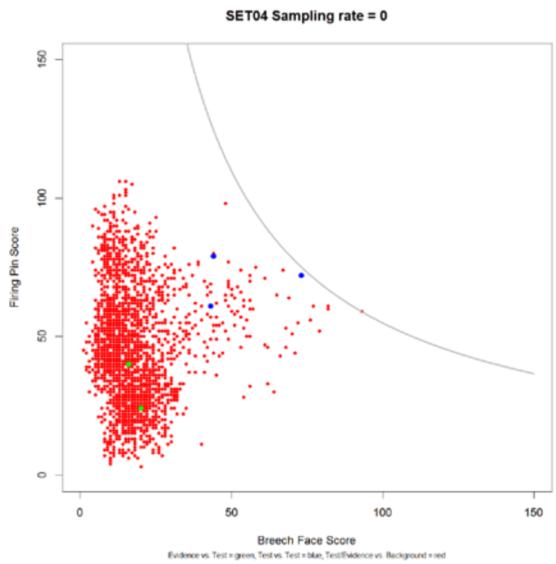

Figure 190: Firing in versus breech face scores for Set 04

Figure 190 represents the data for Set 04. The test-versus-test scores are situated just below the best-known non-match line although well outside the main clustering of the background data. The two test-versus-evidence scores lie well within the main cluster of the background samples and well below best-known non-match line. These results would support a non-match between the evidence and test samples.

Set 05

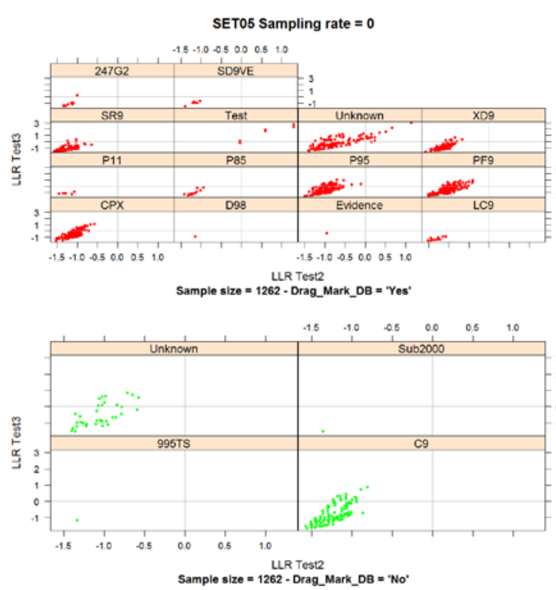

Figure 191: Comparison of the log likelihood ratios (Test2 and Test3) of evidence cartridge case from Set 05

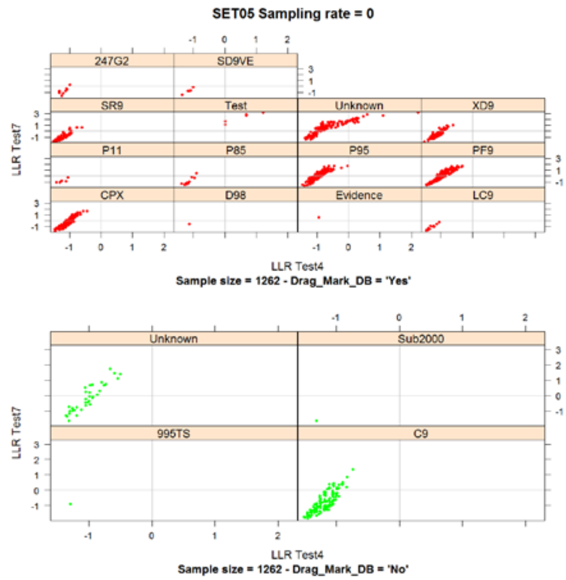

Figure 192: Comparison of the log likelihood ratios (Test4 and Test7) of evidence cartridge case from Set 05

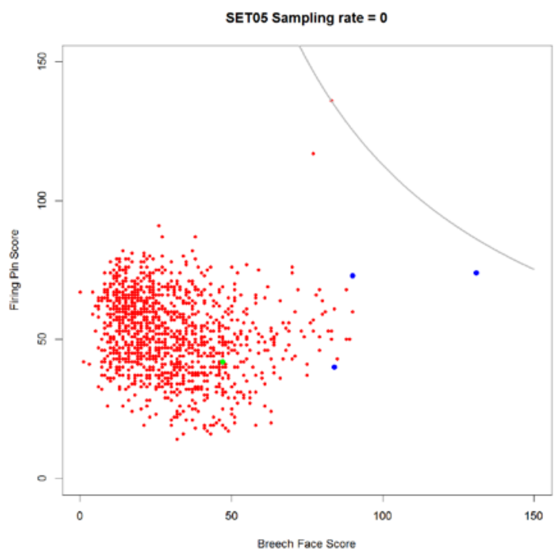

Figure 193: Firing pin versus breech face scores for Set 05

Figure 193 represents the data from Set 05. One of the test-versus-test results lies below the best-known non-match line, however, it has a significantly higher breech face score than all the other background data. The other two test-versus-test results are at the maximum breech face score periphery of the background data. Two of the three test-versus-test results lie at the maximum periphery of the firing pin scores. The single returned evidence-versus-test score lies within the bulk of the background data. These data support a non-match result.

Set 06

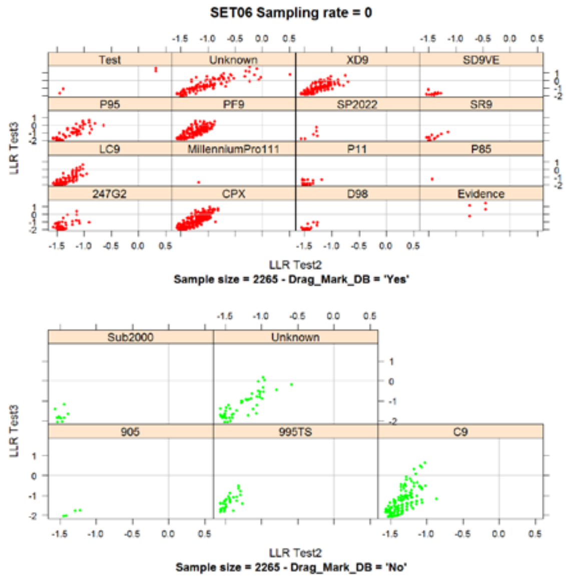

Figure 194: Comparison of the log likelihood ratios (Test2 and Test3) of evidence cartridge case from Set 06

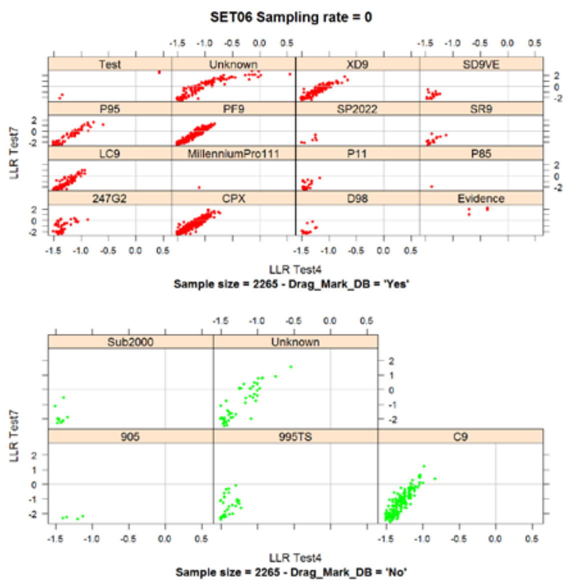

Figure 195: Comparison of the log likelihood ratios (Test4 and Test7) of evidence cartridge case from Set 06

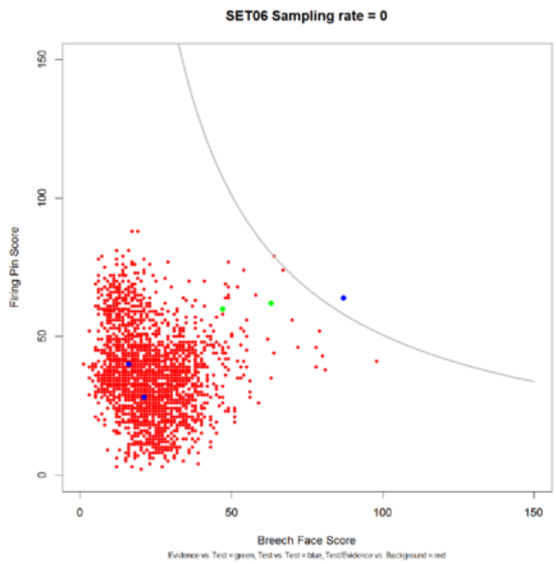

Figure 196: Firing pin versus breech face scores for Set 06

Figure 196 represents the data from Set 06. One of the test-versus-test scores lies slightly above the best-known non-match. The other two lie well within the bulk of the background data. The two test-versus-evidence scores lie outside the main cluster of the background data but below the best-known non-match curve. Given the AFTE theory of identification, these must be considered a non-match.

Set 07

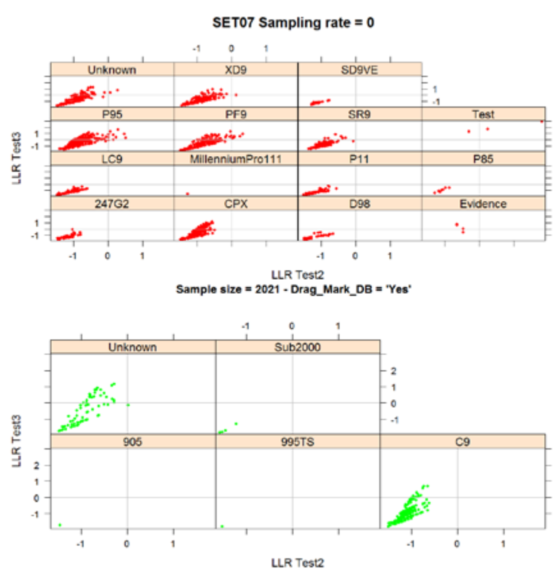

Figure 197: Comparison of the log likelihood ratios (Test2 and Test3) of evidence cartridge case from Set 07

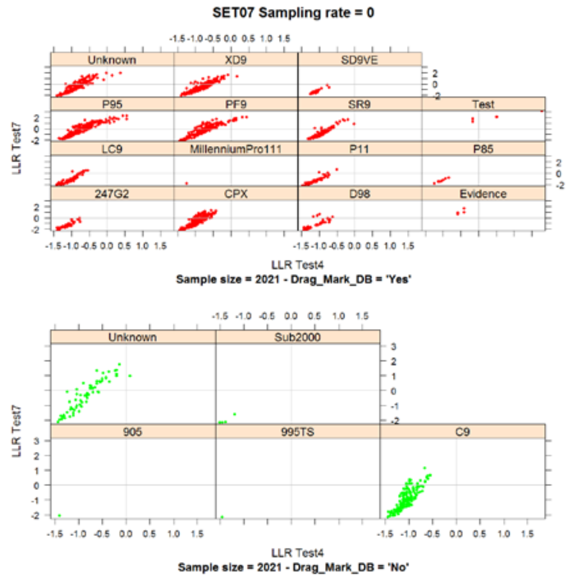

Figure 198: Comparison of the log likelihood ratios (Test4 and Test7) of evidence cartridge case from Set 07

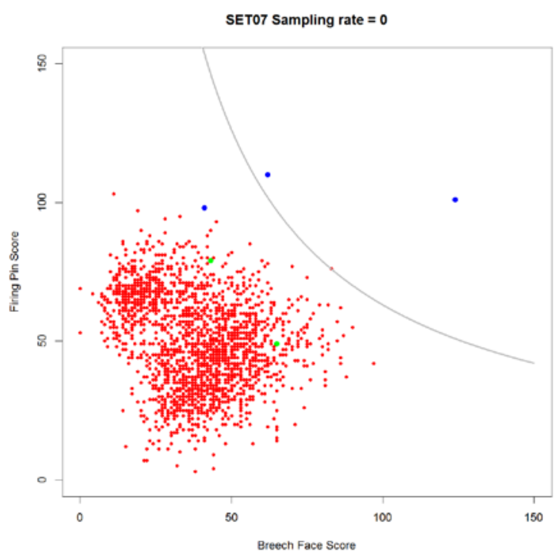

Figure 199: Firing pin versus breech face scores for Set 07

Figure 199 represents data of set 07. Two of the three test-versus-test data lie well above the best-known non-match. The third test-versus-test result, although under the curve, lies at the upper boundary of firing pin scores. All of the firing pin scores of these three points are of the same order. The two test-versus-evidence points lie within the bulk of the background data and well below the best-known non-match could. Consequently a match cannot be called between the test and evidence cartridge cases.

Set 08

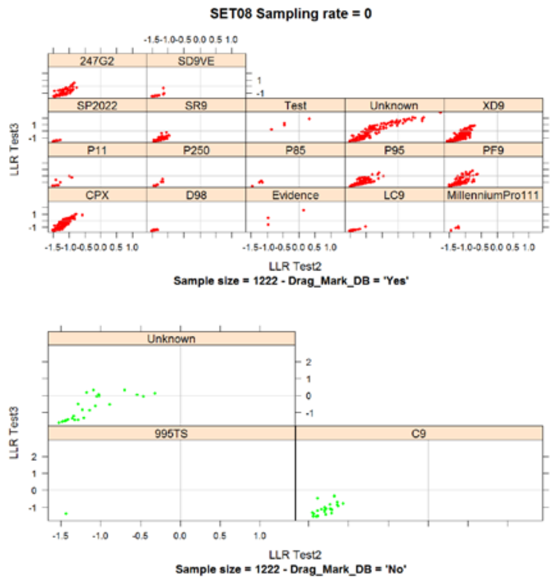

Figure 200: Comparison of the log likelihood ratios (Test2 and Test3) of evidence cartridge case from Set 08

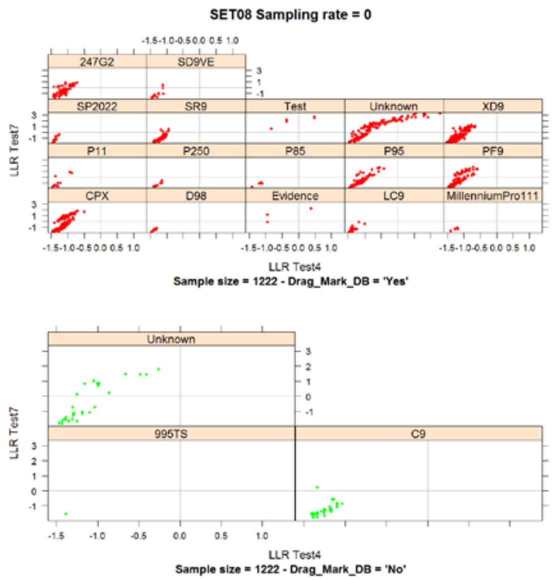

Figure 201: Comparison of the log likelihood ratios (Test4 and Test7) of evidence cartridge case from Set 08

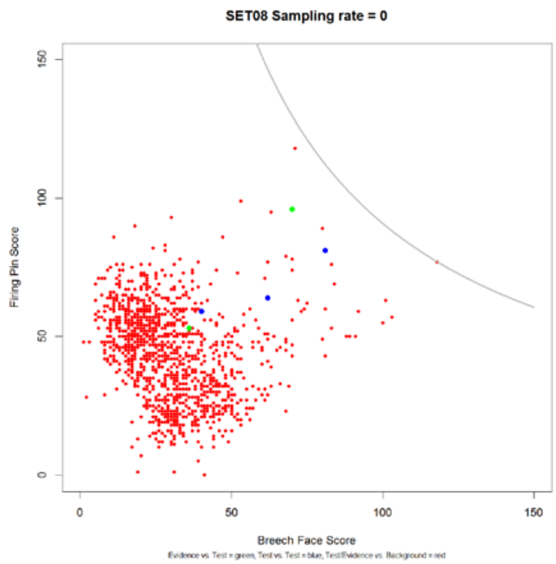

Figure 202: Firing pin versus breech face scores for Set 08

Figure 202 represents the data from Set 08. The test-versus-test scores are clustered together at the outer periphery of background data well below the best-known non-match. One of the test-versus-evidence data lies at the outer edge of the background data. The other is well within the bulk of the background data. These scores must be interpreted as a non-match.

Set 09

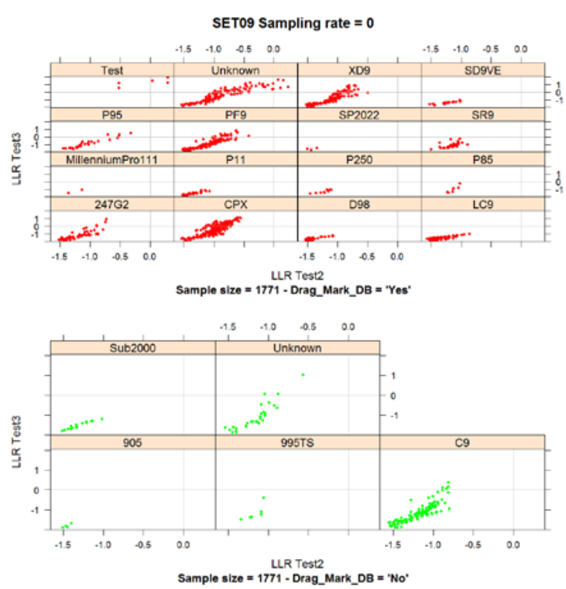

Figure 203: Comparison of the log likelihood ratios (Test2 and Test3) of evidence cartridge case from Set 09

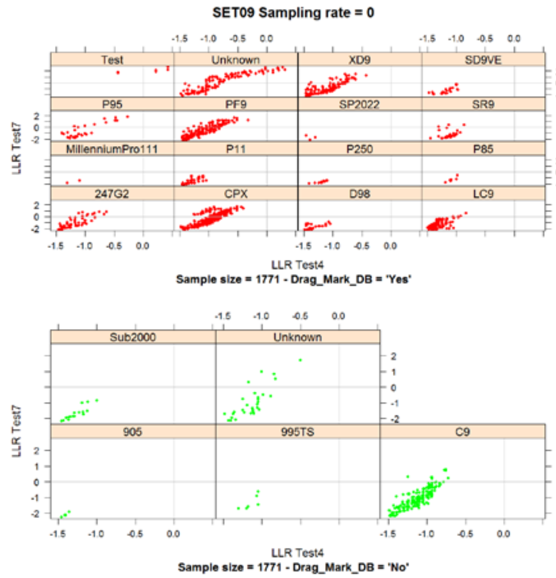

Figure 204: Comparison of the log likelihood ratios (Test4 and Test7) of evidence cartridge case from Set 09

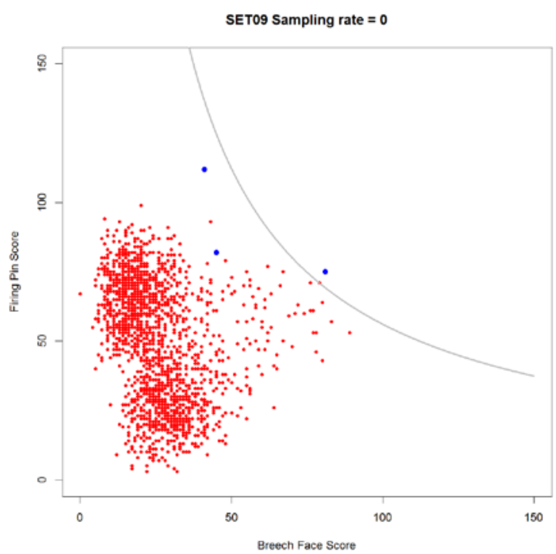

Figure 205: Firing pin versus breech face scores for Set 09

Figure 205 represents the data of Set 09. Test-versus-test data for the set lie around the best-known non-match curve, with one point the above it and another having the maximum firing pin score. No results were returned from IBIS for the question sample. This implies that questioned versus test scores were worse than any of the scores represented in this plot. A non-match is inferred.

Set 10

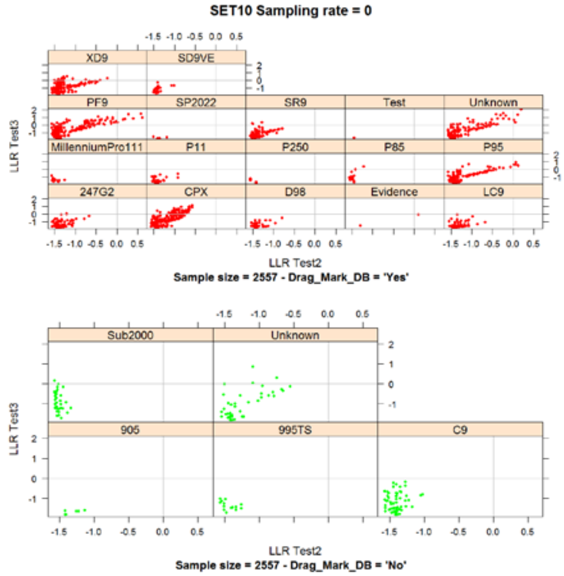

Figure 206: Comparison of the log likelihood ratios (Test2 and Test3) of evidence cartridge case from Set 10

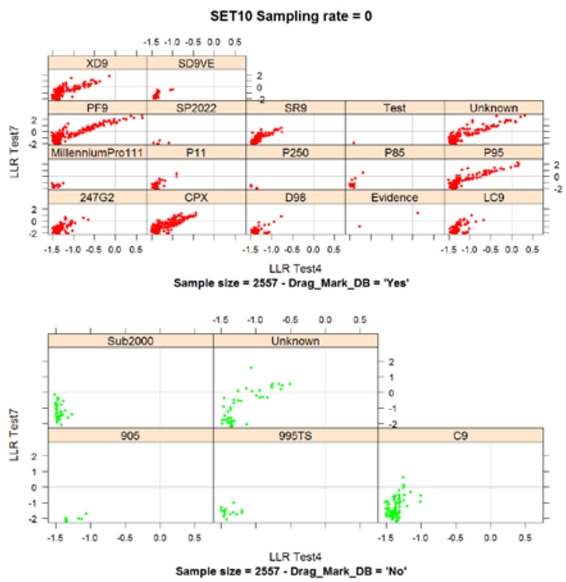

Figure 207: Comparison of the log likelihood ratios (Test4 and Test7) of evidence cartridge case from Set 10

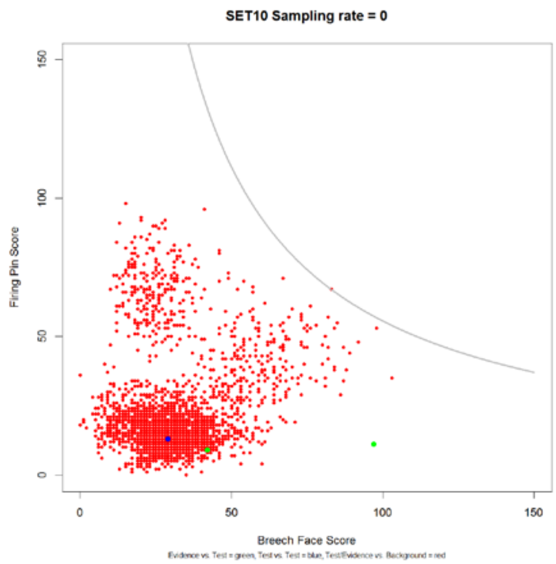

Figure 208: Firing pin versus breech face scores for Set 10

Figure 208 represents the data for Set 10. The background data in this sample set appears to consist of three clusters, one dense cluster close to scores of about 25 for both firing pin and breech face. One of the test-versus-test scores and one of the test-versus-evidence results lie within this cluster. A second test-versus-evidence score is found at a relatively high breech face score but with a firing pin score with a similar value as that of the main background cluster. These data suggest a non-match.

Set 11

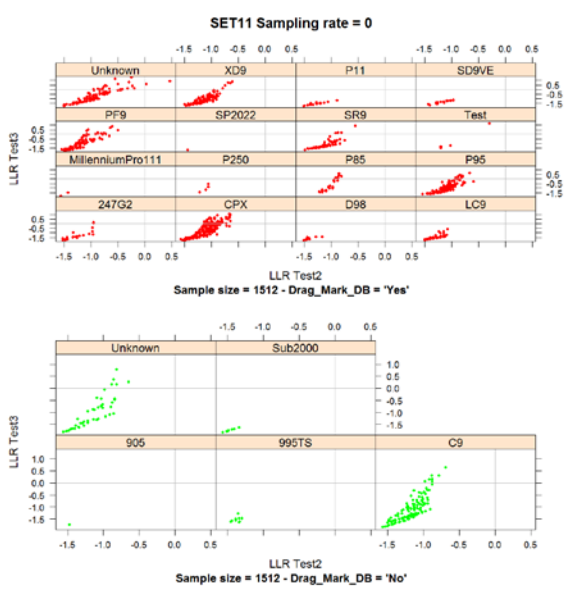

Figure 209: Comparison of the log likelihood ratios (Test2 and Test3) of evidence cartridge case from Set 11

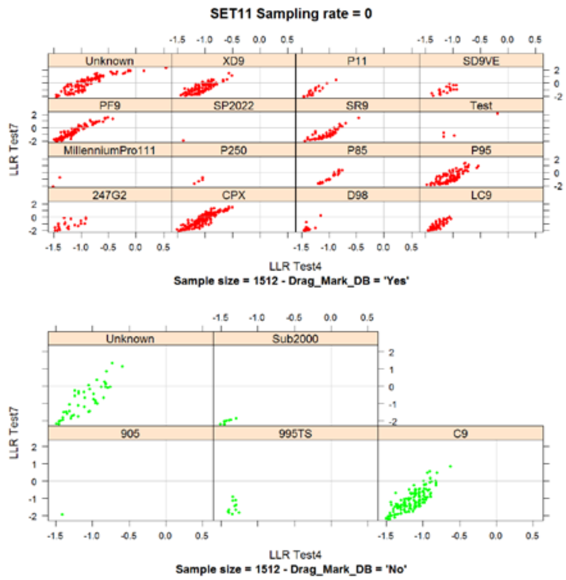

Figure 210: Comparison of the log likelihood ratios (Test4 and Test7) of evidence cartridge case from Set 11

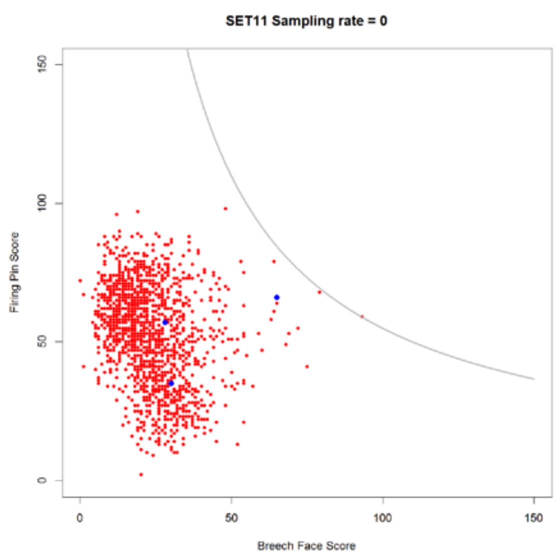

Figure 211: Firing pin versus breech face scores for Set 11

Figure 211 represents the data for Set 11. The test-versus-test scores lie well below the best-known non-match curve. Two of these points are well within the background cluster. No test-versus-evidence scores were returned. This is considered a non-match.

Set 12

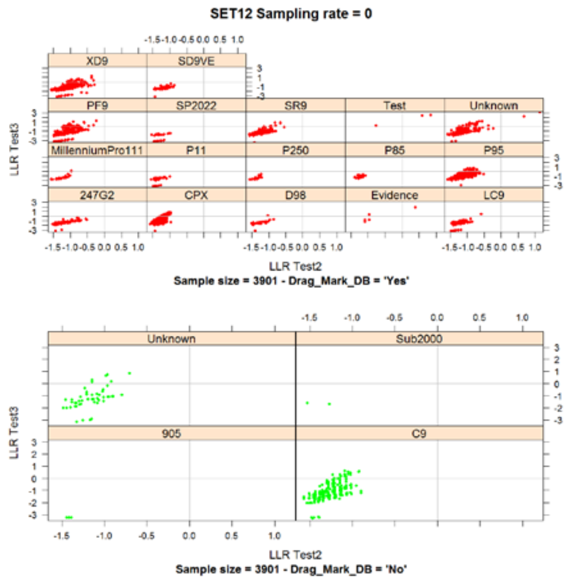

Figure 212: Comparison of the log likelihood ratios (Test2 and Test3) of evidence cartridge case from Set 12

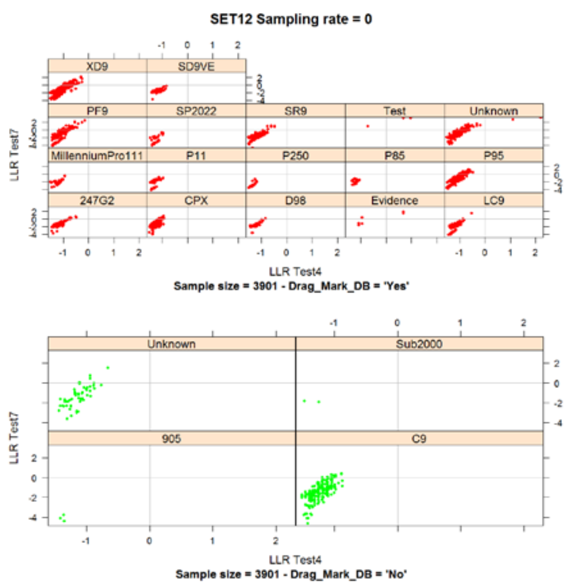

Figure 213: Comparison of the log likelihood ratios (Test4 and Test7) of evidence cartridge case from Set 12

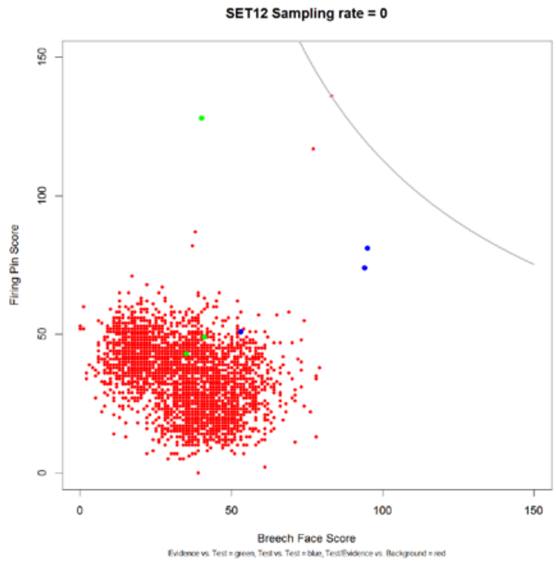

Figure 214: Firing pin versus breech face scores for Set 12

Figure 214 represents the data from Set 12. Two test-versus-evidence scores and one test-versus-test score are found in the main cluster of the background data. Two test-versus-test scores are at higher breech face and firing pin scores but below the best-known non-match curve. A third test-versus-evidence score is at a relatively high firing pin score well below the best-known non-match curve. These data support a non-match.

Set 13

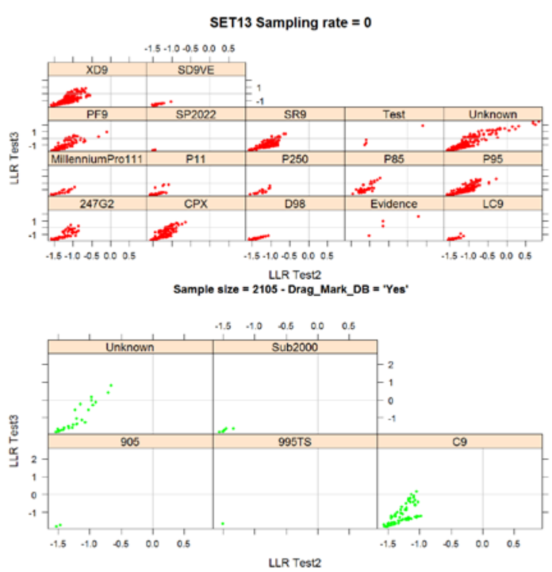

Figure 215: Comparison of the log likelihood ratios (Test2 and Test3) of evidence cartridge case from Set 13

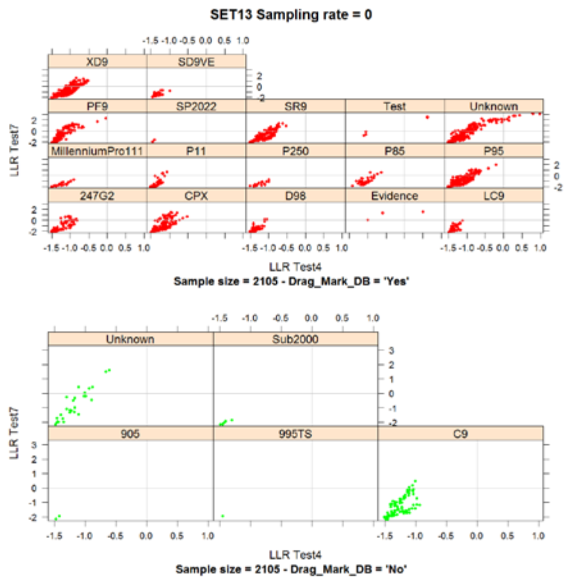

Figure 216: Comparison of the log likelihood ratios (Test4 and Test7) of evidence cartridge case from Set 13

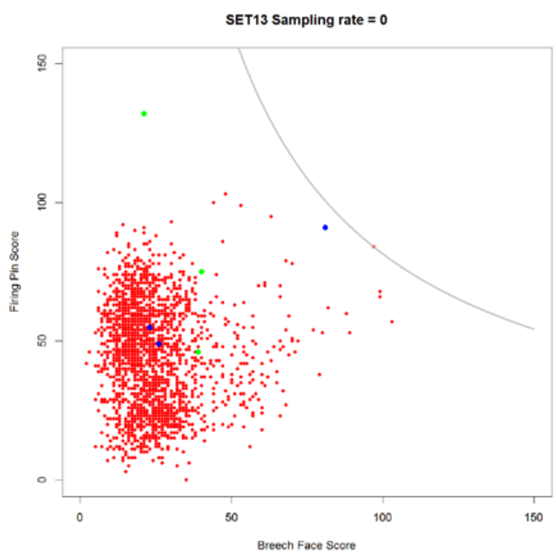

Figure 217: Firing pin versus breech face scores for Set 13

Figure 217 represents the data of Set 13. Two test-versus-test scores and two test-versus-evidence scores lie within the background data cluster. A single test-versus-test score lies just below the best-known non-match curve, and one test-versus-evidence score lies below the best-known non-match curve but with a very high firing pin score (the highest score of the data set). These data may support a match between the test and evidence cartridge cases.

Set 14

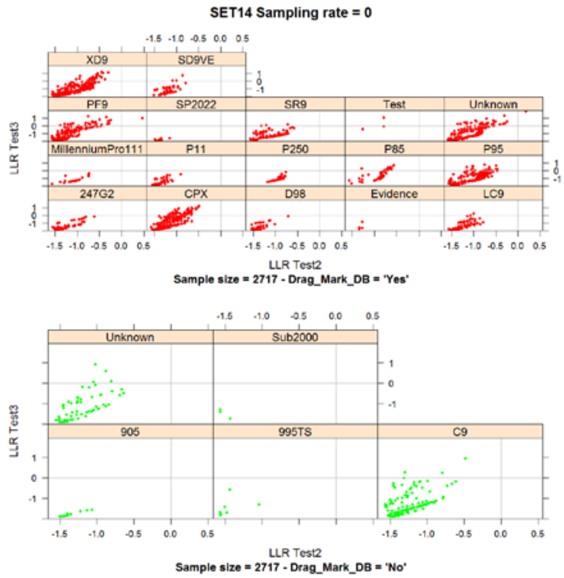

Figure 218: Comparison of the log likelihood ratios (Test2 and Test3) of evidence cartridge case from Set 14

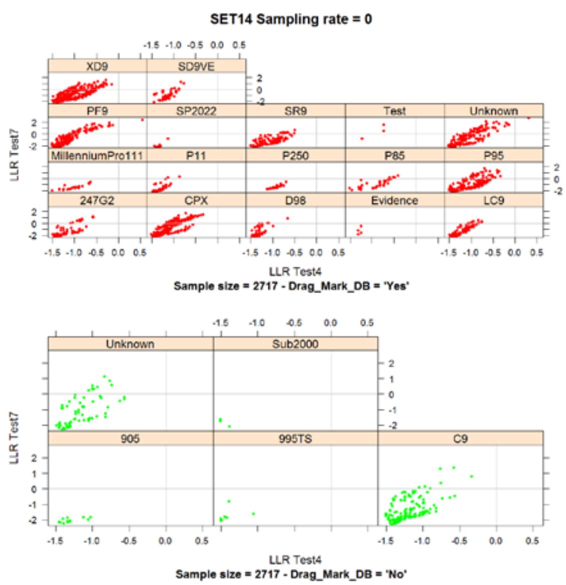

Figure 219: Comparison of the log likelihood ratios (Test4 and Test7) of evidence cartridge case from Set 14

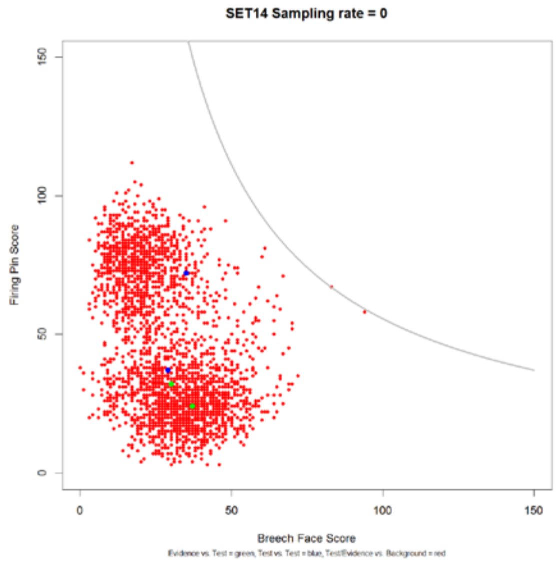

Figure 220: Firing pin versus breech face scores for Set 14

Figure 220 represents the data for Set 14. All of the test-versus-test and evidence-versus-test scores lie well within the background cluster and well below best-known non-match curve. This supports a finding of a non-match.

Set 15

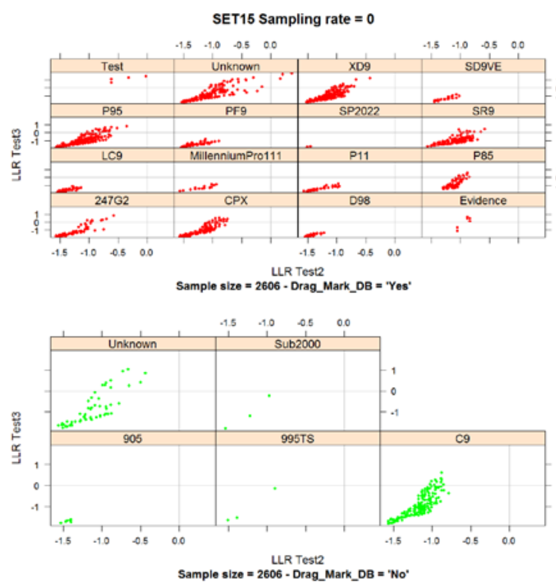

Figure 221: Comparison of the log likelihood ratios (Test2 and Test3) of evidence cartridge case from Set 15

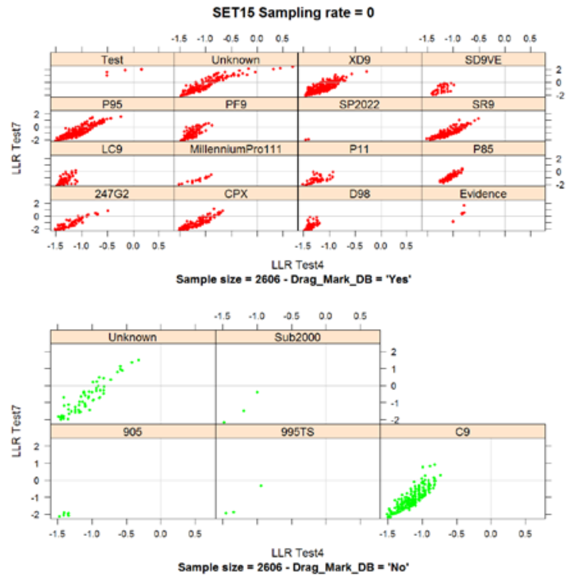

Figure 222: Comparison of the log likelihood ratios (Test4 and Test7) of evidence cartridge case from Set 15

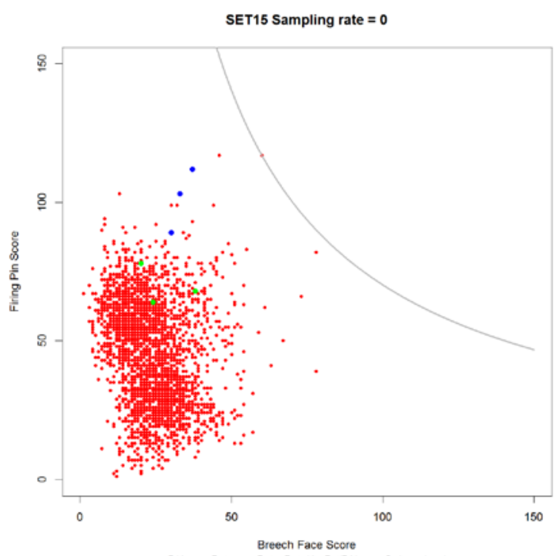

Figure 223: Firing pin versus breech face scores for Set 15

Figure 223 represents the data of Set 15. The test-versus-test scores are clustered at relatively high firing pin scores, whilst the test-versus-evidence scores lie at the high end of the firing pin score in the background cluster. All of these scores lie well below the best-known non-match curve and subsequently a finding of a non-match is given.

Set 16

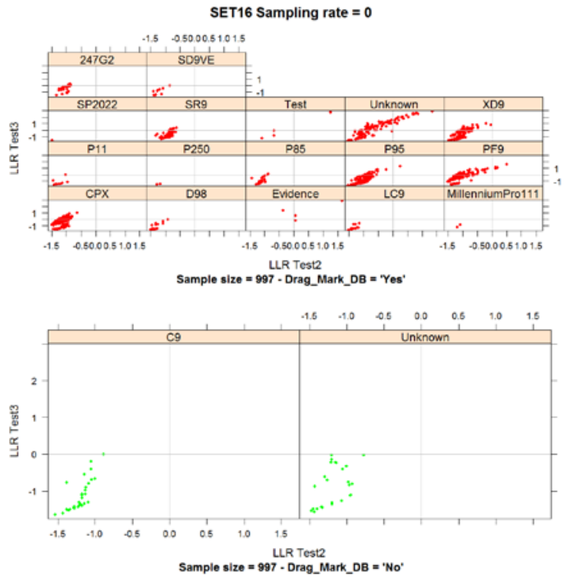

Figure 224: Comparison of the log likelihood ratios (Test2 and Test3) of evidence cartridge case from Set 16

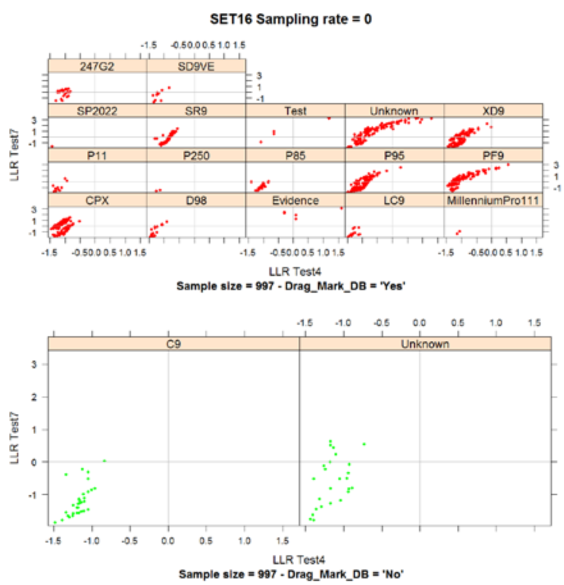

Figure 225: Comparison of the log likelihood ratios (Test4 and Test7) of evidence cartridge case from Set 16

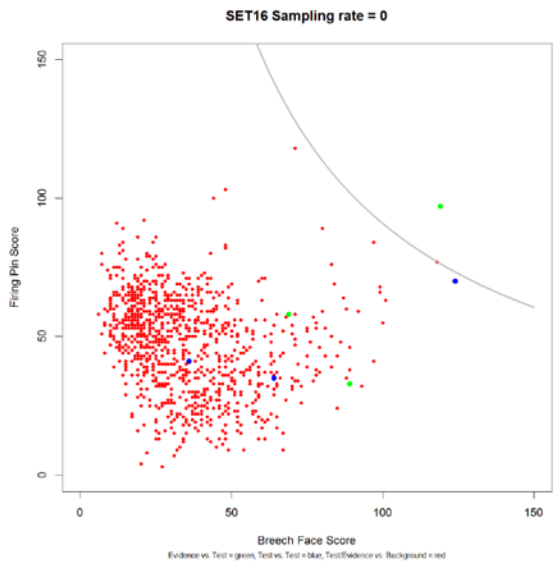

Figure 226: Firing pin versus breech face scores for Set 16

Figure 226 represents the data of Set 16. This data set returned a relatively small number of scores for the background cluster. One of the test-versus-test scores lies just below the best-known non-match curve, and one of the test-versus-evidence scores lies above the curve. The remaining four scores fall within the background cluster. The position of the single question versus test score supports a finding of a match.

Set 17

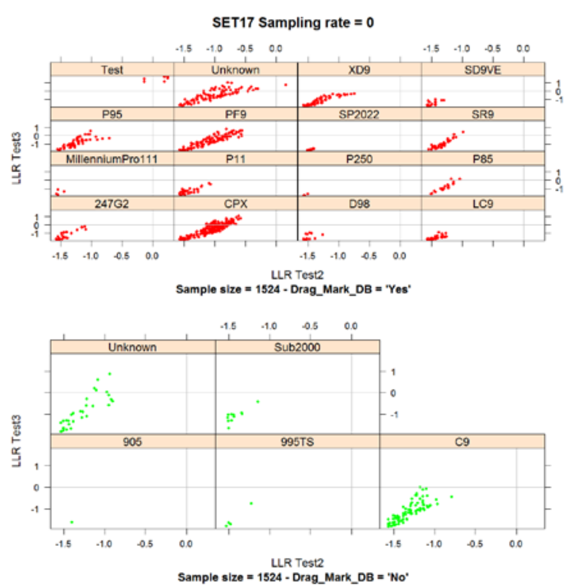

Figure 227: Comparison of the log likelihood ratios (Test2 and Test3) of evidence cartridge case from Set 17

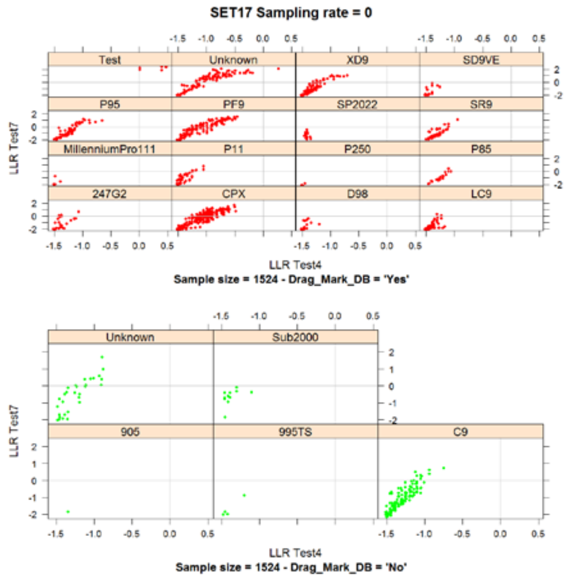

Figure 228: Comparison of the log likelihood ratios (Test4 and Test7) of evidence cartridge case from Set 17

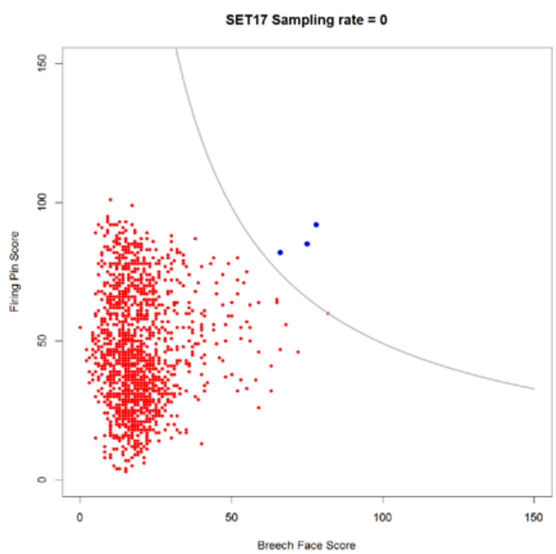

Figure 229: Firing pin versus breech face scores for Set 17

Figure 229 represent the data for Set 17. In this data set all three of the test-versus-test scores lie closely clustered above the best-known non-match curve. This indicates that these three cartridge cases seem to have very similar characteristics. There are, however, no test-versus-evidence scores indicating that this is a non-match.

Set 18

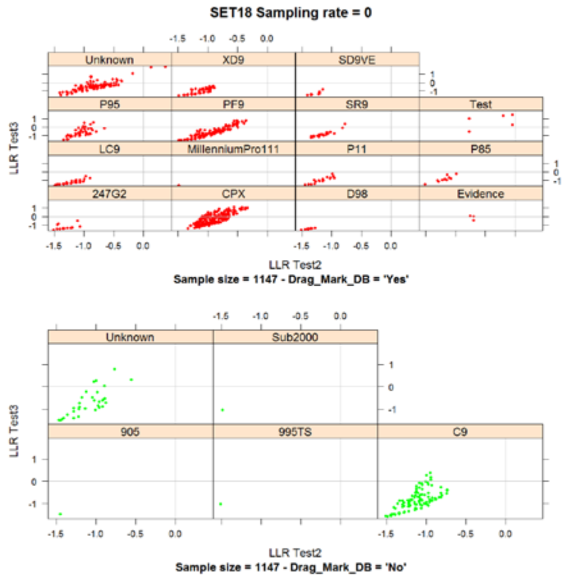

Figure 230: Comparison of the log likelihood ratios (Test2 and Test3) of evidence cartridge case from Set 18

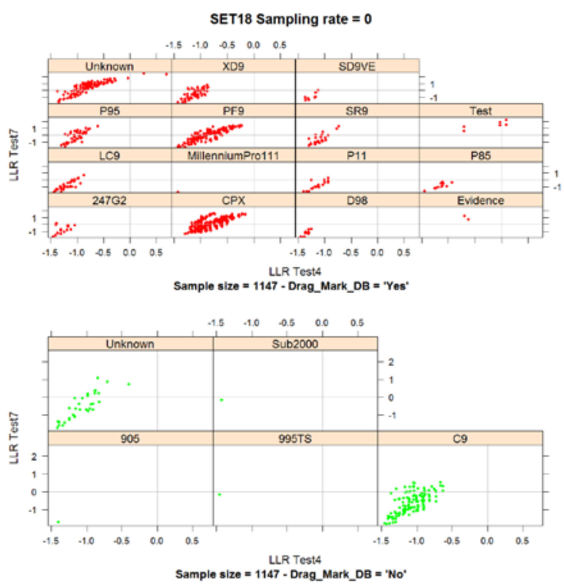

Figure 231: Comparison of the log likelihood ratios (Test4 and Test7) of evidence cartridge case from Set 18

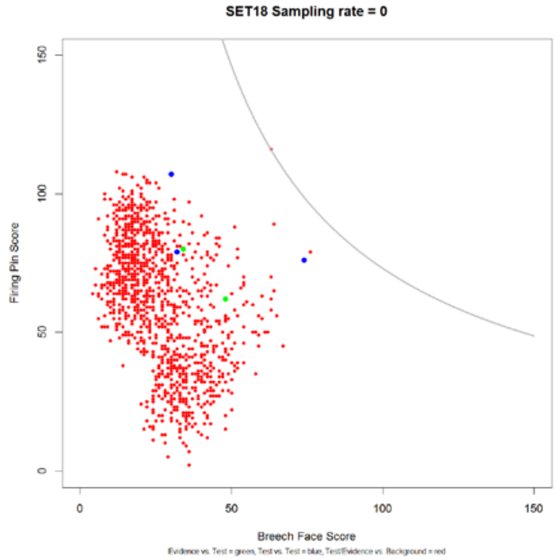

Figure 232: Firing pin versus breech face scores for Set 18

Figure 232 represents the data of Set 18. One of the test-versus-test scores lies just below the best-known non-match. The other test-versus-test scores and the test-versus-evidence scores all lie well within the background cluster. These data support a finding of non-match.

Set 19

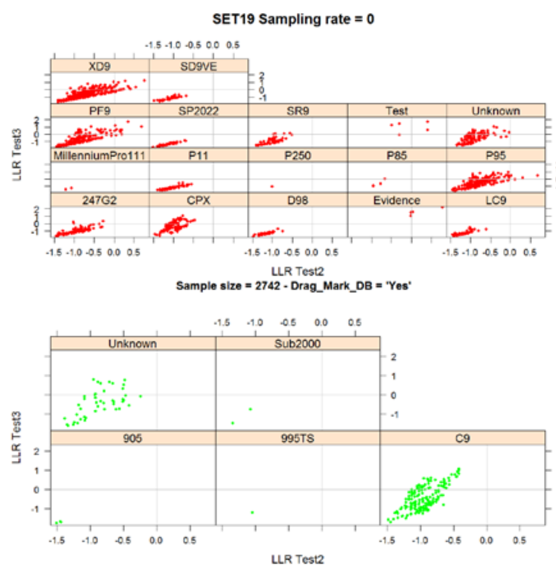

Figure 233: Comparison of the log likelihood ratios (Test2 and Test3) of evidence cartridge case from Set 19

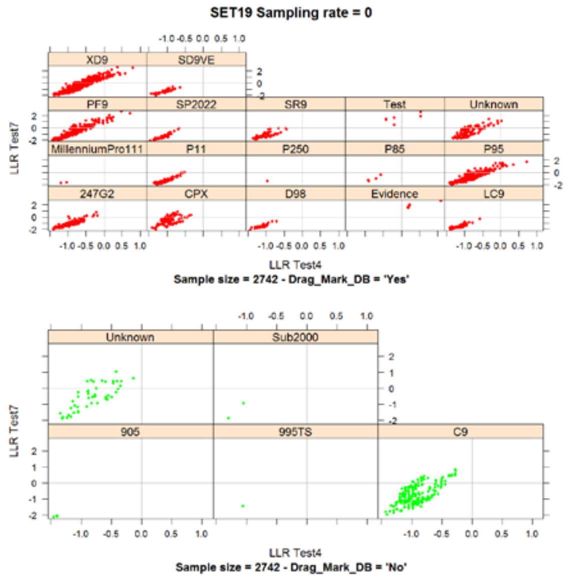

Figure 234: Comparison of the log likelihood ratios (Test4 and Test7) of evidence cartridge case from Set 19

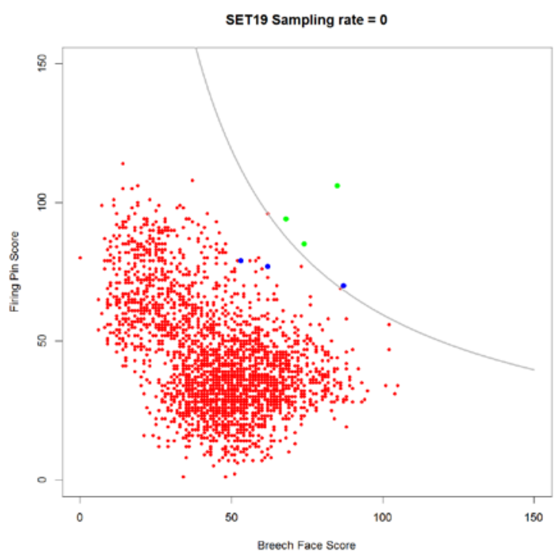

Figure 235: Firing pin versus breech face scores for Set 19

Figure 235 represents the data from Set 19. The test-versus-test scores are well clustered but straddle the best-known non-match curve. The three test-versus-evidence scores are clustered just above the best-known non-match. These data support a finding of a match.

Set 20

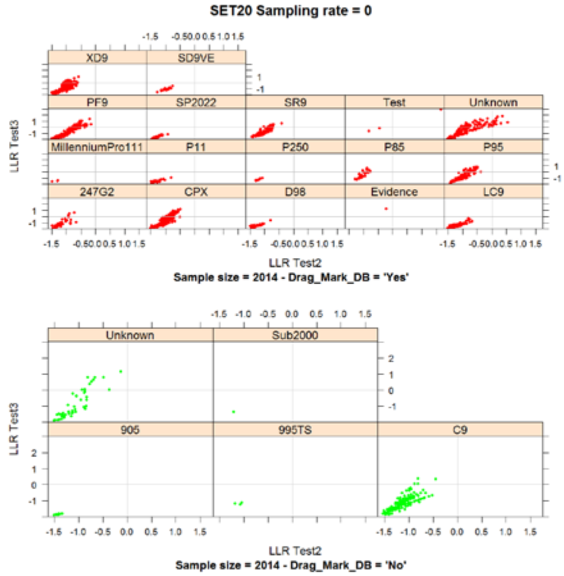

Figure 236: Comparison of the log likelihood ratios (Test2 and Test3) of evidence cartridge case from Set 20

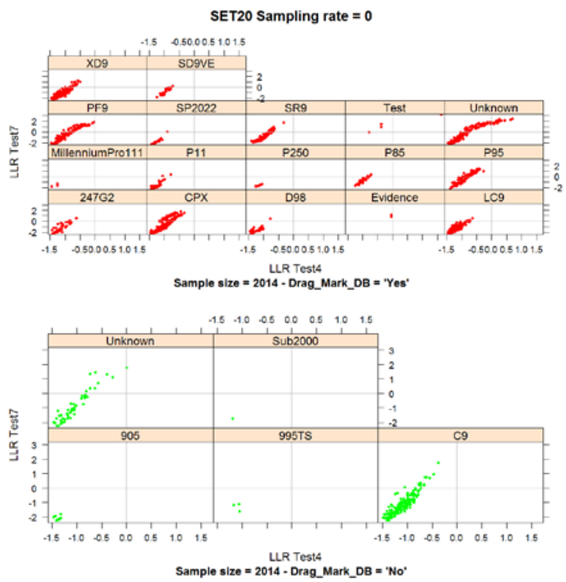

Figure 237: Comparison of the log likelihood ratios (Test4 and Test7) of evidence cartridge case from Set 20

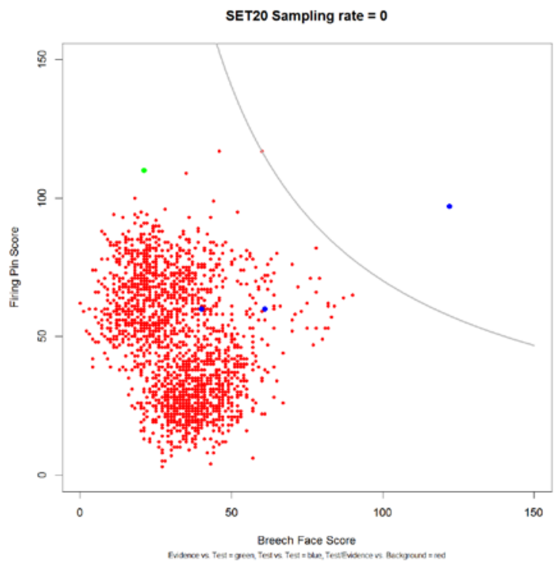

Figure 238: Firing pin versus breech face scores for Set 20

Figure 238 represents the data of Set 20. Two of the test-versus-test scores lie within the cluster of the background data, and the 3rd lies well beyond the best-known non-match curve. A single test-versus-evidence score lies at high value side of the firing pin scores but below the best-known non-match curve. Given the firing pin score, this data may weakly support a call of a non-match.

Baldwin test set: Likelihood ratio analysis

This data set comprises of 20 subsets each comprising of one question cartridge case and three known cartridge cases. In the previous analysis, all of the results of the cartridge cases from the 25 SR9s used to condition the firearms were removed from the data set. This section includes all of these data. The database results are classified into four categories: Evidence, Tests, SR9Test, and Background. The SR9Test category includes all of the cartridge cases that were used to condition the firearms. The Background category includes all of the firearms (SR9's and other models). The data analysis indicates how the results are conditioned on these categories.

The log-likelihood ratios (LLR) tests used in the study are Test 4 and Test 7. Test 4 uses the type of firing pin (Firing_Pin_Type_Sample), the presence or absence of a drag mark (Drag_Mark_Sample), and the breech face (BF) and firing pin (FP) scores generated by IBIS as evidence in the calculation of the likelihood ratio (LR). Test 7 uses the same evidence as Test 4, but in addition includes both the FP rank (Rank) and the BF rank (Rank_BF). The prior probabilities are assessed based upon the characteristics of each of the known and questioned cartridge cases. The use of these priors will result in different LLRs from those generated by using no information from the known and questioned cartridge cases. All of the results are provided as LLRs. The priors can be assessed by inspection of the Bayesian network.

The file containing the entire test data were run against the Bayesian network to compute the posterior probabilities. These results were then used to calculate the likelihood ratio using the appropriate prior probabilities.

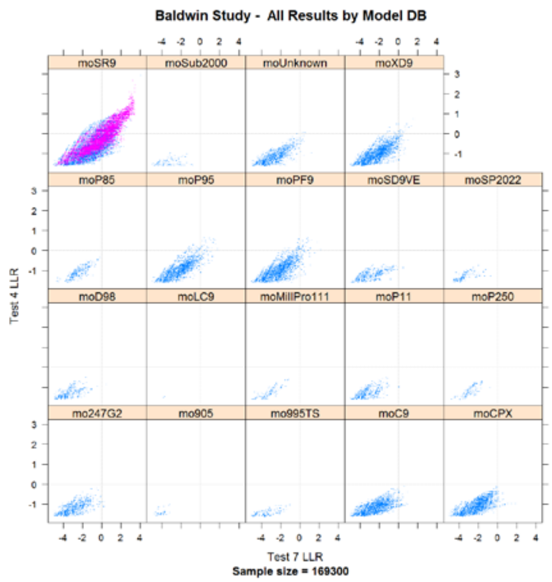

Figure 239: Baldwin data – LLR (Test 4) vs. LLR (Test7) by model DB

Figure 239 demonstrates the LLRs (Test 4 and Test 7) by firearm models in database. It is assumed that all of the known (Test) and questioned (Evidence) samples originate from a Ruger SR9 pistol. Figure 239 also includes lines indicating LLRs of zero (LLR=0 implies that the evidence is neutral) which help to interpret the impact of the LLRs. For all of the models other than SR9, the LLRs are either close to or less than 0. These results support the proposition that the cartridge case was fired by a different firearm. For the SR9's, they are a number of instances where non-matches have a LLR greater than zero.

Figure 240 indicates the match and non-match results of the evidence in the data set. It is evident that most of the non-matches (blue dots) for the evidence against the test samples at LLRs less than zero. They are a number of matches (pink dots) that also have a LLR of less than zero.

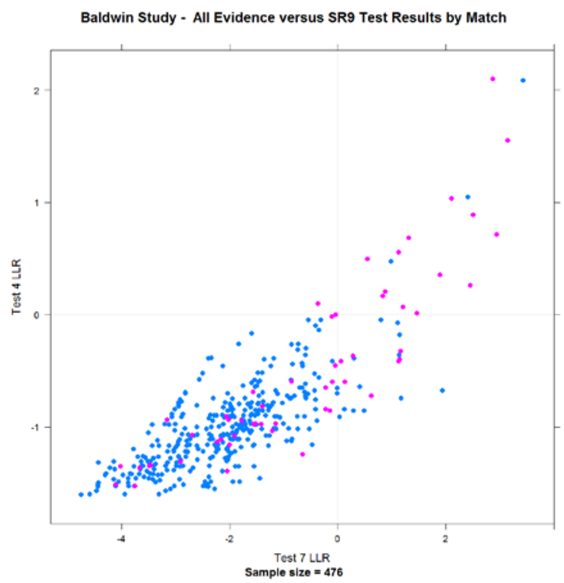

Figure 240: All results for Evidence-versus-test by match (LLR(Test4) vs LLR(Test7))

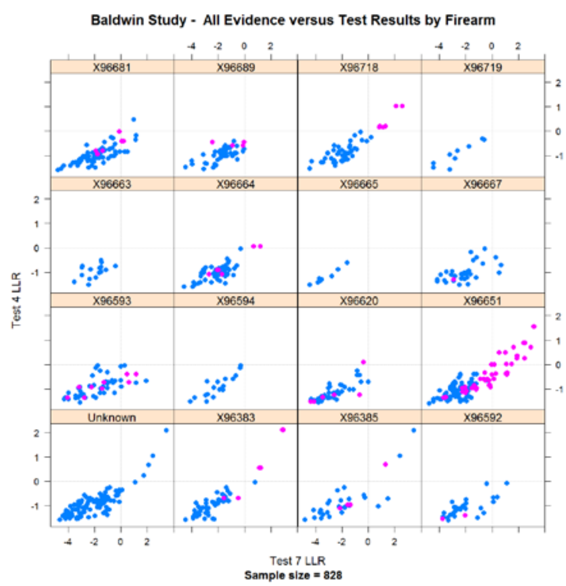

Figure 241: All results for Evidence vs. Test by SR9 firearms

Figure 241 indicates the results by the serial numbers of the database SR9's. Firearm X96651 has a large number of results since it was used in three of the twenty tests. In the Unknown firearm, four results have high LLR values. These are comparisons between two of the Questioned cartridge cases belonging to elimination sets (SET05-Q1: SET12-K2, SET12-K3 (X96385) and SET11-Q1: SET18-K2, SET18-K3).

Table 46: Association of likelihood ratio with verbal equivalent (Evett & Buckleton)

LLR of Evidence C	**Conclusion**
LLR = 0	The evidence is neutral
0 < LLR <= 1	The evidence slightly supports C
1 < LLR <= 2	The evidence supports C
2 < LLR <= 3	The evidence strongly supports C
3 < LLR	The evidence very strongly supports C

The verbal scales for LLRs are given in

Table 46. These are then applied in Figure 242 and compared to the Truth and Baldwin results. These data are given as follows: each question sample per set is associated with the LLRs of each test and each known cartridge case responding to a search on IBIS. The columns entitled "Evidence…" are the verbal scales associated with the LLR in the preceding column. These should be read as "The evidence _____ supports *sgp*/*dgp*". The "*same gun proposition*" (*spg*) and "*different gun proposition*" (*dpg*) are abbreviated for brevity. The cells highlighted in light green indicate that the LLR is in support of the Truth-value. Those in pink do not support the Truth-value. The empty cells provide strong support either for or against the Truth-value as per their color (dark green or dark red). It should be noted that some questioned samples have LLRs both in support and against the Truth-value indicating the variability of the results.

Q Sample	Baldwin	Truth	K1: LLR 4	Evidence…	K2: LLR 4	Evidence…	K3: LLR 4	Evidence…	K1: LLR 7	Evidence…	K2: LLR 7	Evidence…	K3: LLR 7	Evidence…
SET01-Q1	Incon	ID	-1.08	dgp	0.07	slightly sgp	-0.91	slightly dgp	-2.70	strongly dgp	1.21	sgp	-2.06	strongly dgp
SET02-Q1	FN	ID			-1.31	dgp			-2.91	strongly dgp				
SET03-Q1	Incon	ID	-0.69	slightly dgp	2.10	strongly sgp	0.56	slightly sgp	-1.56	dgp	2.87	strongly sgp	1.12	sgp
SET04-Q1	Incon	ID			-1.39	dgp	-1.53	dgp			-2.05	strongly dgp	-3.75	very strongly dgp
SET05-Q1	FP	Elim					-0.98	slightly dgp					-1.51	dgp
SET06-Q1	Incon	ID	-0.72	slightly dgp	-0.40	slightly dgp			0.62	slightly sgp	1.15	sgp		
SET07-Q1	Incon	ID	-0.45	slightly dgp			-0.60	slightly dgp	-0.05	slightly dgp			-0.10	slightly dgp
SET08-Q1	Incon	ID			-0.97	slightly dgp	0.35	slightly sgp			-1.52	dgp	1.90	sgp
SET09-Q1	FP	Elim												
SET10-Q1	Incon	ID	0.10	slightly sgp			-1.34	dgp	-0.36	slightly dgp			-3.46	very strongly dgp
SET11-Q1	FP	Elim												
SET12-Q1	Incon	ID	-1.11	dgp	-0.98	slightly dgp	0.68	slightly sgp	-2.16	strongly dgp	-1.55	dgp	1.31	sgp
SET13-Q1	Incon	ID	-1.03	dgp	-0.60	slightly dgp	0.49	slightly sgp	-1.21	dgp	0.13	slightly sgp	0.55	slightly sgp
SET14-Q1	FP	Elim			-1.42	dgp	-1.31	dgp			-3.47	very strongly dgp	-3.52	very strongly dgp
SET15-Q1	FP	Elim	-0.81	slightly dgp	-1.02	dgp	-0.81	slightly dgp	-1.23	dgp	-2.42	strongly dgp	-1.38	dgp
SET16-Q1	FN	ID	0.00	is neutral	-0.33	slightly dgp	1.55	sgp	-0.04	slightly dgp	1.17	sgp	3.15	very strongly sgp
SET17-Q1	FP	Elim												
SET18-Q1	FP	Elim					-0.61	slightly dgp					-1.59	dgp
SET19-Q1	Incon	ID	0.17	slightly sgp	0.20	slightly sgp	1.03	sgp	0.83	slightly sgp	0.87	slightly sgp	2.11	strongly sgp
SET20-Q1	FN	ID			-0.02	slightly dgp					-0.11	slightly dgp		

Figure 242: LLR results and Verbal scales[64]

Once the data for the conditioning study are added back into the data set, there are more test results by type. These results are given in Figure 243. These results clearly indicate the improvement of the LLRs for the matching data. This underlines that the variation in the markings are better represented through an increased sample size when IBIS is used as the measuring instrument.

[64] The cells highlighted in light green indicate that the LLR is in support of the Truth-value. Those in pink do not support the Truth-value. The empty cells provide strong support either for or against the Truth-value as per their color (dark green or dark red).

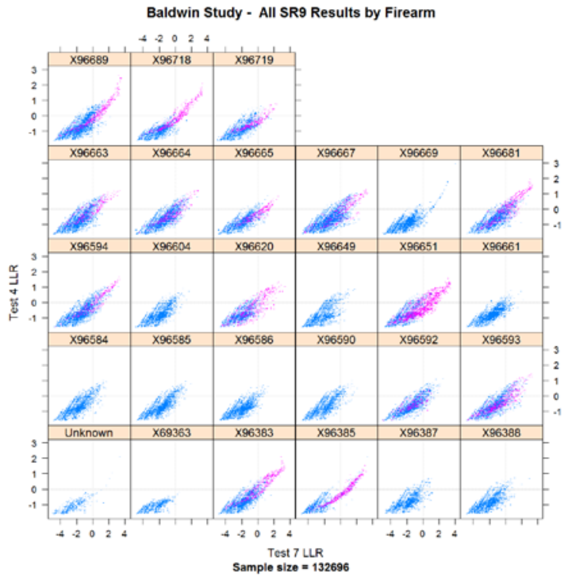

Figure 243: All results for SR9 (including condition data) by firearm

These results are given by each set. In Figure 244 and Figure 246 the results of 8 sets are given.

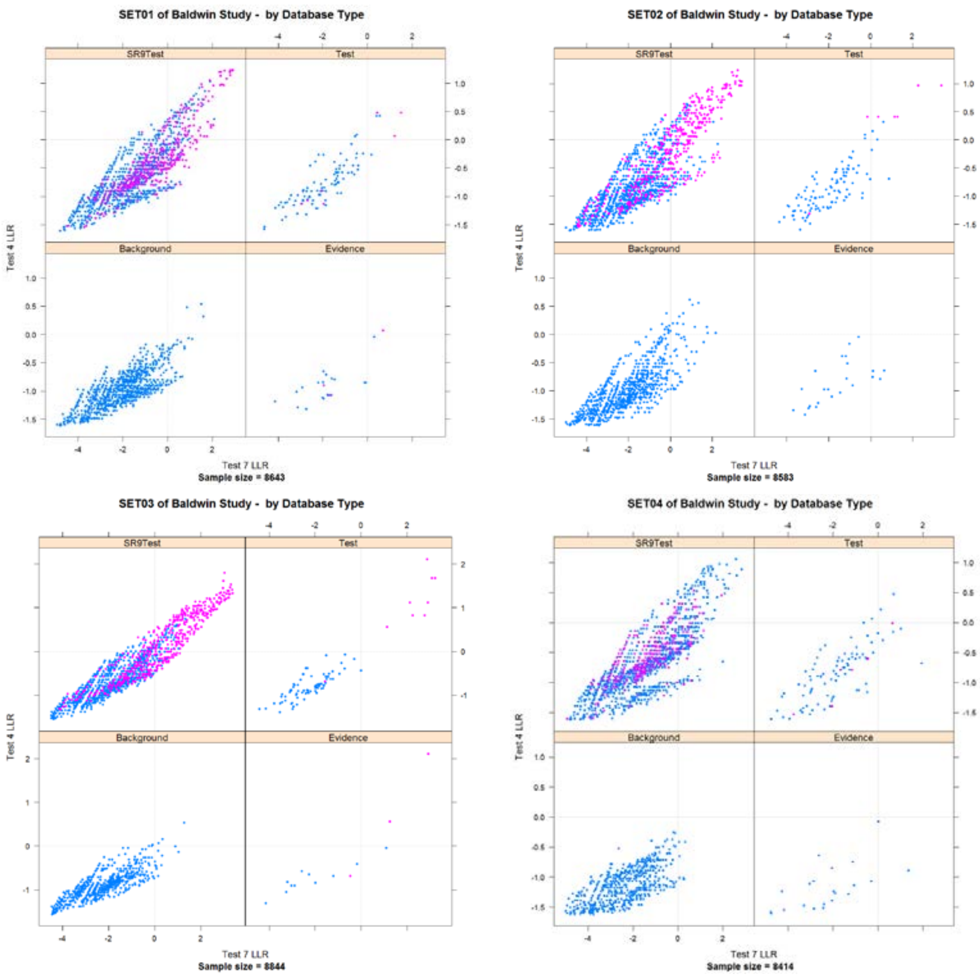

Figure 244: LLRs by DB Type Set 01, Set 02, Set 03, and Set 04

In Figure 245, the results of Set 05 are given. It is noticeable that there is a large number of non-matches with LLRs significantly greater than zero. If sample Q1 of Set 05 is an elimination, then the true identity of the questioned sample is unknown. When considering Test 7 LLRs for this cartridge case, there are only two that do not come from SR9 with serial number X96385. These two are from the SR9 with serial number X96667. If this question sample does, in fact, originate from X96385, then they are 193 matches that are marked as non-matches (blue instead of pink).

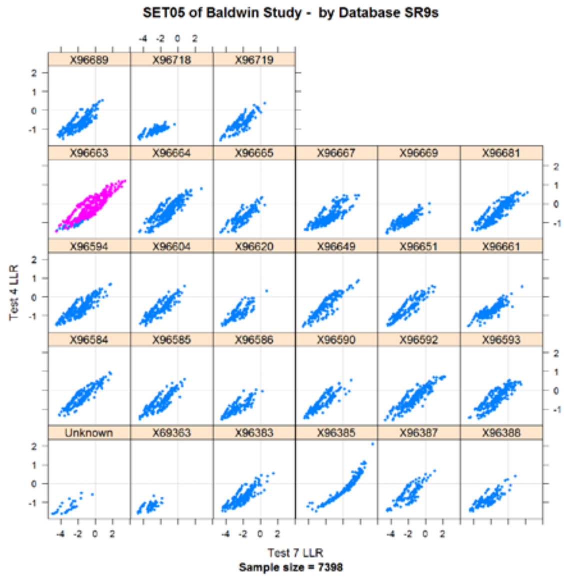

Figure 245: Set 05 Results against SR9's by serial number of DB firearm

The belief that the questioned sample is from Set 05 is supported by the data in Figure 245. For all of the firearms the non-match data extends only slightly beyond the LLRs = 0. For firearm X96385, the LLRs extend to values larger than those of the matches of firearm X96663. The test samples Set 05 originate from firearm X96663.

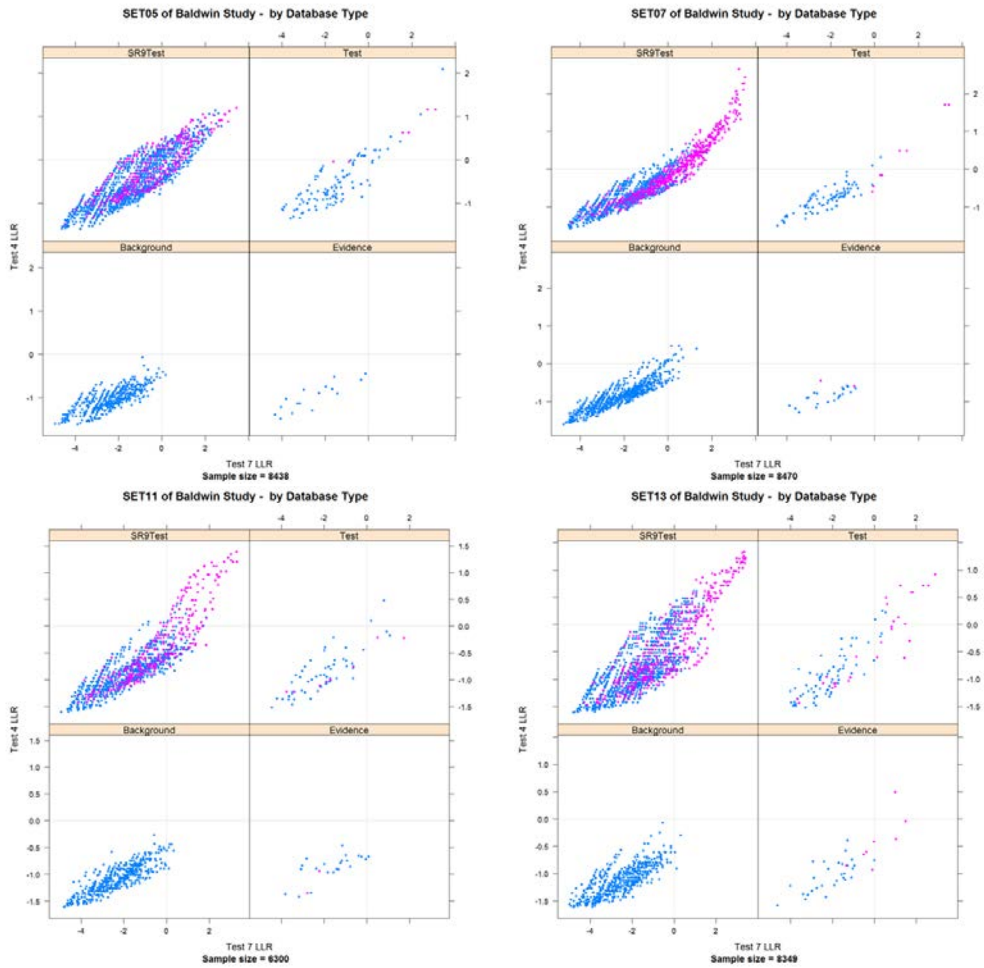

Figure 246: LLRs by DB Type Set 05, Set 07, Set 11, and Set 13

When considering the truth values, Table 47 provides the results of the assessment of the truth values using the LLR (Test 7) values. The top LLRs were used to assess which SR9 firearm was most prevalent. If the firearm was the same as that provided in the truth data, the column labeled LLR (Test 7) Ability was given a checkmark. If there was no specific firearm prevalent in the top values, then approximately equal sign (≈) was placed in the LLR (Test 7) Ability column (inconclusive). For the elimination truth-values (different gun), if a specific firearm was always in the top values, then the serial number of that firearm was placed in the column alongside the checkmark. In the case of set 11 there were only 3 results against everything except the background data. Given this performance, it is postulated that the firearm that fired the questioned cartridge case is not part of the original 25 SR9s used in the study.

Table 47: Results with full SR9 data. Eliminations include most probable firearm which fired the questioned cartridge case

Set	Letter	Serial Number	Truth	Baldwin Results	LLR(Test 7) Ability
Set 01	D3	X96664	Same Gun	Inconclusive	≈
Set 02	D5	X96667	Same Gun	False Negative	✓
Set 03	A1	X96383	Same Gun	Inconclusive	✓
Set 04	B5	X96592	Same Gun	Inconclusive	≈
Set 05	D2	X96663	Different Gun	False Positive	✓ (X96385)
Set 06	B5	X96593	Same Gun	Inconclusive	≈
Set 07	E3	X96689	Same Gun	Inconclusive	✓
Set 08	C5	X96651	Same Gun	Inconclusive	✓
Set 09	C1	X96594	Different Gun	False Positive	✓ (X96719)
Set 10	C3	X96620	Same Gun	Inconclusive	≈
Set 11	B5	X96593	Different Gun	False Positive	? (Firearm not in DB)
Set 12	A2	X96385	Same Gun	Inconclusive	✓
Set 13	C5	X96651	Same Gun	Inconclusive	✓
Set 14	C3	X96620	Different Gun	False Positive	✓ (X96669)
Set 15	E2	X96681	Different Gun	False Positive	✓ (X96590)
Set 16	C5	X96651	Same Gun	False Negative	✓
Set 17	E5	X96719	Different Gun	False Positive	✓ (X96593)
Set 18	D4	X96665	Different Gun	False Positive	✓ (X96383)
Set 19	E4	X96718	Same Gun	Inconclusive	✓
Set 20	E2	X96681	Same Gun	False Negative	✓

Conclusion

Test and "evidence" samples from another DFSC study ("Baldwin study") were provided. These represented sample sets examined by the firearms examiner in a "black box" type study. The data were handled in two situations. Since there were approximately 200 cartridge cases of each firearm used in the Baldwin study in the database, the comparisons were run with these data both excluded and included.

Excluded background: The ground truth data and the firearms examiner test results were provided. For all of the cases where examiners made an inconclusive determination the truth was that the cartridge case was fired from the same gun as the test fires (same-gun). For the examiner false negatives Set 02 and Set 20 agreed with the examiner results. For Set 01, the correct result was achieved with LLR (Test 7) being better than LLR (Test 4).

In Sets 05, 09, 11, 14, 15, 17, and 18 the firearms examiners made false positive attributions. Out of the 20 comparisons, there were eight true positives, seven true negatives, five false negatives and zero false positives. In all instances of eliminations, the support for the different-gun hypothesis was, at minimum, strong.

When the full dataset was used the LLR (Test 7) had difficulty with Sets 01, 04, 06, and 10. For the eliminations, another candidate firearm (from the 25 firearms) tested was identified as the source of the unknown cartridge case. For Set 11, the evidence cartridge case was identified as being from a firearm outside of the original test set but the test cartridge cases were fired by a pistol within the test set.

Bayesian network website

WVU has conducted extensive research and data analysis on various firearms, including cartridge case comparisons. One of the best ways to describe data is by fitting it to a statistical model. Bayesian statistics offers an approach with a natural framework to deal with parameter and model uncertainty. The end goal of Bayesian analysis is to provide a distribution for the knowledge gained (i.e. what was learned) about the parameter from the data. Netica™, a Norsys™ Software Corp program, is a simple, reliable, and high performing Bayesian network development software. A Bayesian network is a model that reflects the states of the given population being modeled and describes how those states are related by probabilities. The aim of this chapter is to provide an easy to follow user manual for setting up and utilizing the Netica™-based cartridge case individualization web interface.

The first step of this manual is deployment, making the Bayesian network of cartridge case individualization available for use. The developed web interface can be hosted on Apache Tomcat server version 6. In order to deploy it on a server follow the instructions (Figure 247):

1. Copy and extract the archive file of the source code of the web interface.
2. Open http://127.0.0.1:8080/ in a browser i.e. open the home screen of the Tomcat server.
3. Navigate to the "Tomcat Manager". Typically opening the http://127.0.0.1:8080/manager/html should take to the Tomcat Manager.
4. Scroll down till the "Deploy" section.
5. In the "Context Path" write: "/Netica" (forward slash is necessary)
6. In the "WAR or Directory URL" write the path to the extracted folder "Netica" of the source folder provided.
7. Click "Deploy".
8. If a message "OK - Deployed application at context path /Netica" appears the web interface is hosted successfully.
9. A restart of the Apache server (depending on your server configuration) may be required before starting to utilize the web interface.

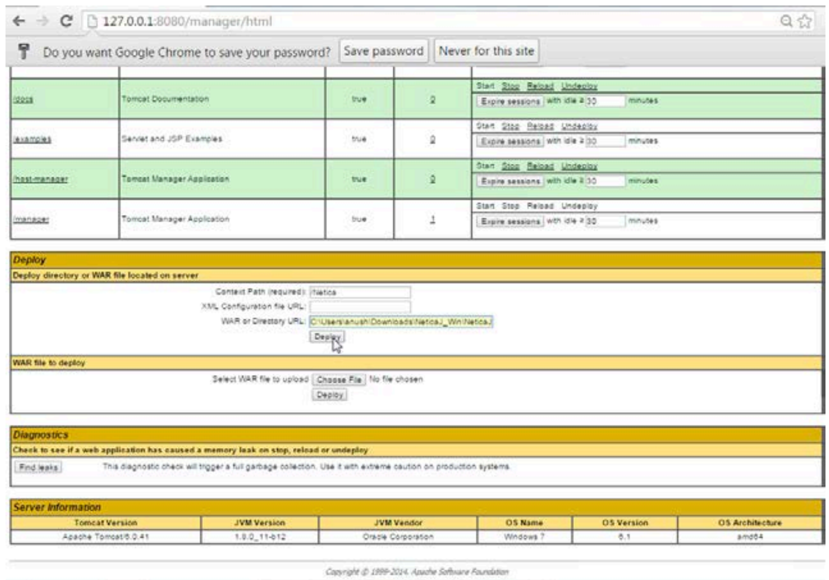

Figure 247: Illustration of deployment steps four through seven of the Netica ™ based web interface

There are various scenarios for which the web interface can be utilized. Three specific cases were chosen to highlight to the user.

Case 1 utilizes the breech face (BF), firing pin (FP), and ranks from the IBIS system scores to find the match probability and likelihood ratio values. The "Case 1 Interface" should be accessible at http://127.0.0.1:8080/Netica/doInterface9mm.jsp.

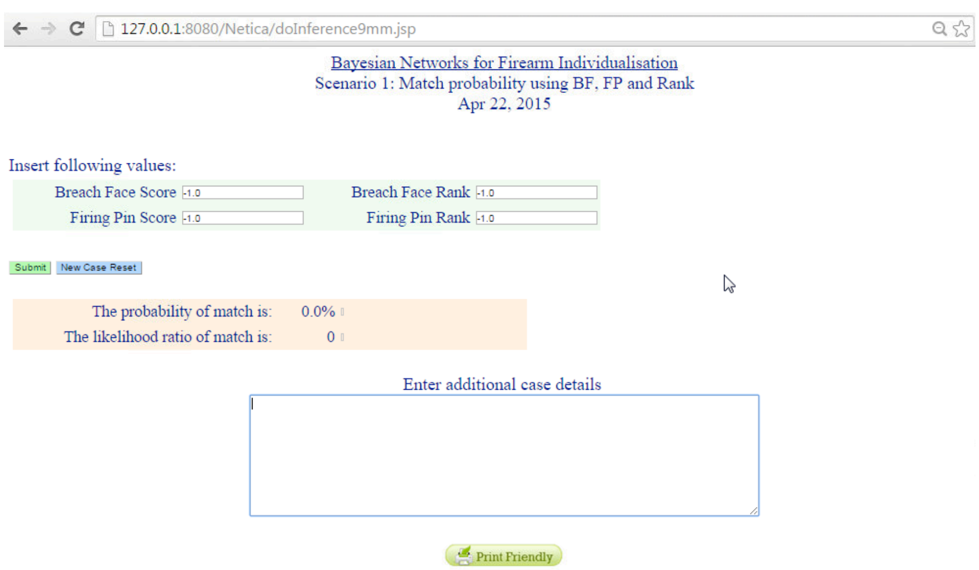

Figure 248: Case 1 web interface display after link is first accessed

The user can then input the BF, FP, BF Rank, and FP Rank as obtained from the IBIS system into the respective fields (Figure 248).

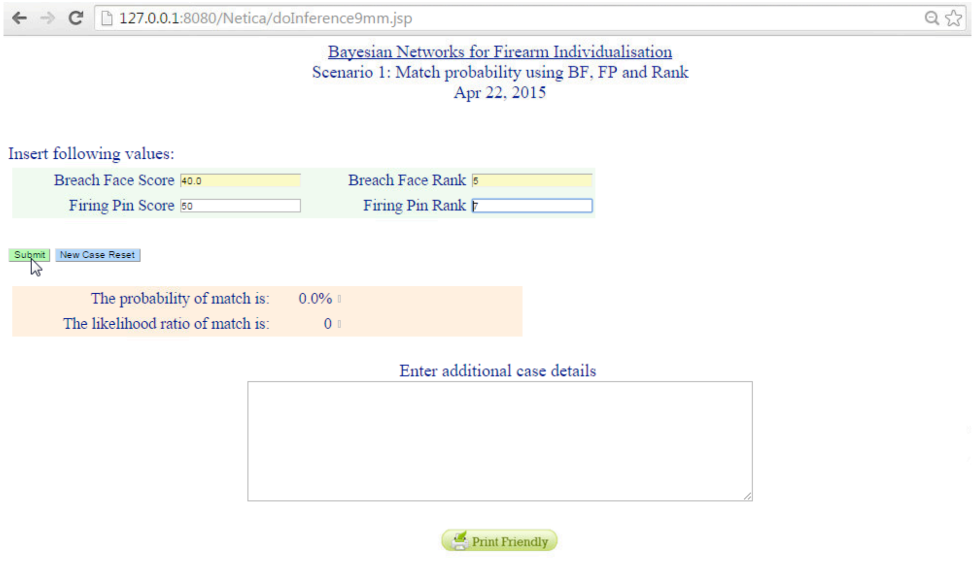

Figure 249: Example of IBIS system scores entered in the appropriate areas

For Case 1, the BF score was entered as 40.0, the FP score as 50, the BF Rank as five, and the FP Rank as seven. After inputting all the fields, click "Submit," the green box with the mouse arrow over it in Figure 249.

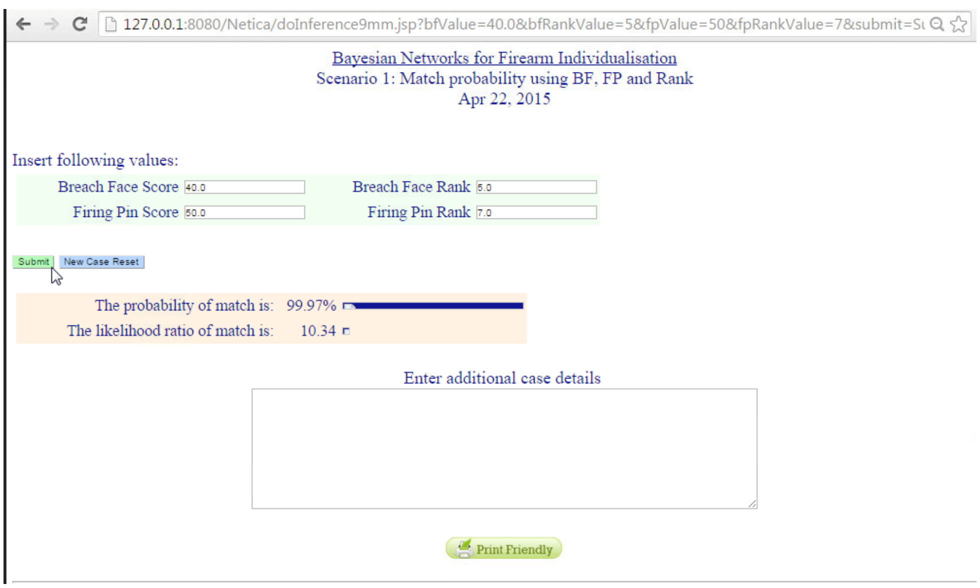

Figure 250: Calculated match probability and likelihood ratio of Case 1 data input.

On clicking the “Submit” button, the “probability of match” and “likelihood ratio of match” should appear. In Figure 250, the resulting probability of match returned at the value of 99.97% (P(Match=Yes|E)) and the likelihood ratio of a match returned at the value of 10.34.

Table 48: Standards for numerical and verbal expression of likelihood ratios

Recommended likelihood ratio terminology	
Numerical expression	Verbal expression (support)
> 1–10	Weak or limited
10–100	Moderate
100–1,000	Moderately strong
1,000–10,000	Strong
10,000–1,000,000	Very strong
> 1,000,000	Extremely strong

The Association of Forensic Science Providers (UK)[65] put forth standards for the interpretation of likelihood ratios. The value of 10.34 from Figure 250 would return a moderate strength of support that the two cartridge cases being compared in Case 1 would be a match.

There is also an option to add case-specific details to each comparison to allow for better organization (Figure 251).

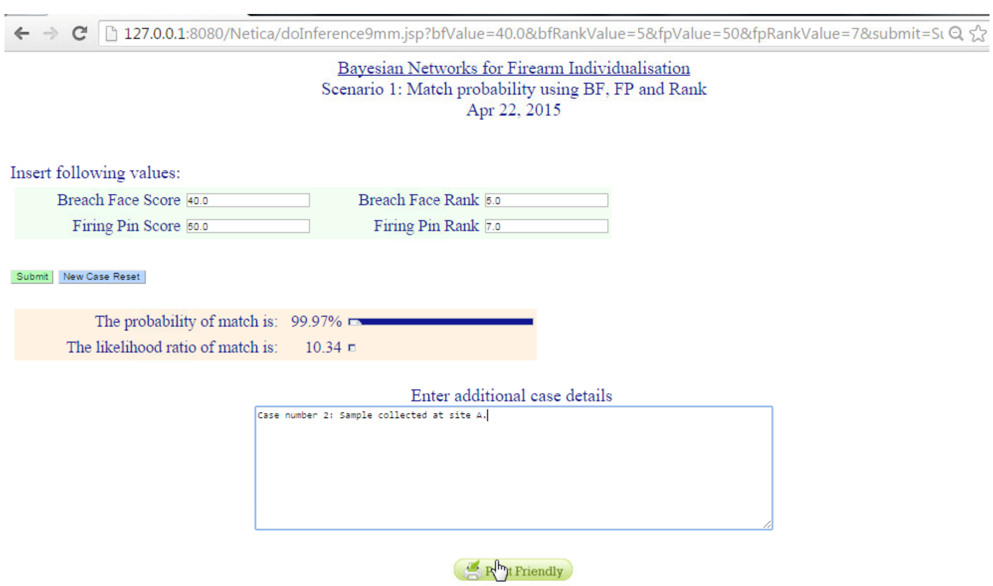

Figure 251: Case-specific details added in the corresponding textbook.

[65] Association of Forensic Science Providers. (2009). Standards for the formulation of evaluative forensic science expert opinion. Science & Justice, 49, 161–164. doi:10.1016/j.scijus.2009.07.004

These pages can be printed to be added to a case file, court documents, personal notes, etc., if needed.

The goal of Case 2 is to predict the best possible match of the make and model of an unknown firearm. This situation could be applicable when there is no firearm recovered, from the scene or persons of interest, but a cartridge case has been collected. The "Case 2 Interface" should be accessible at http://127.0.0.1:8080/Netica/doInterface9mm2.jsp.

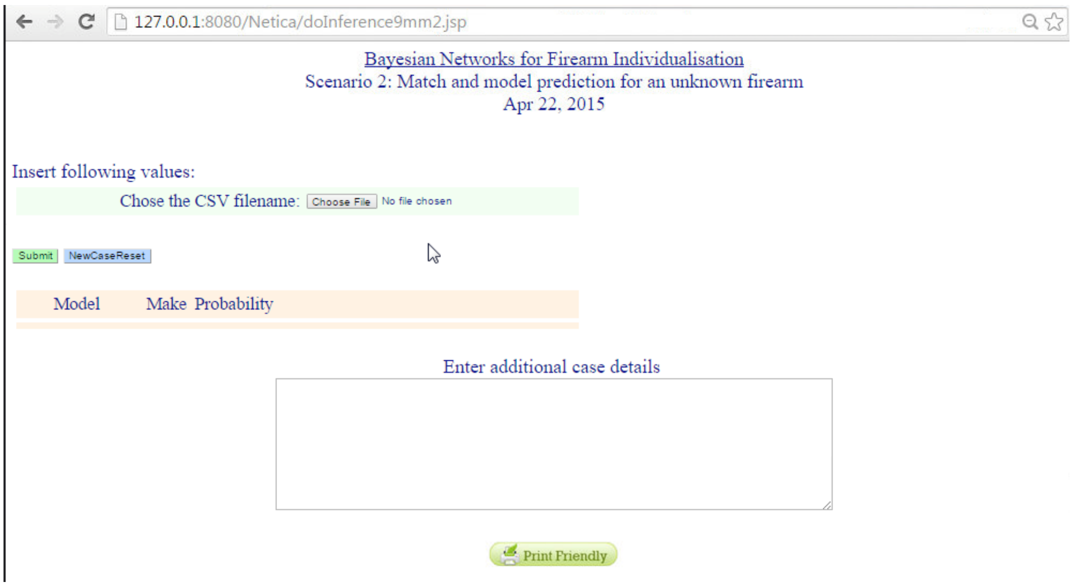

Figure 252: Case 2 web interface display after link is first accessed.

Once the web page is opened and appears as in Figure 252, a *.csv filename must be chosen to insert the values for the prediction. Click on "Choose File" button.

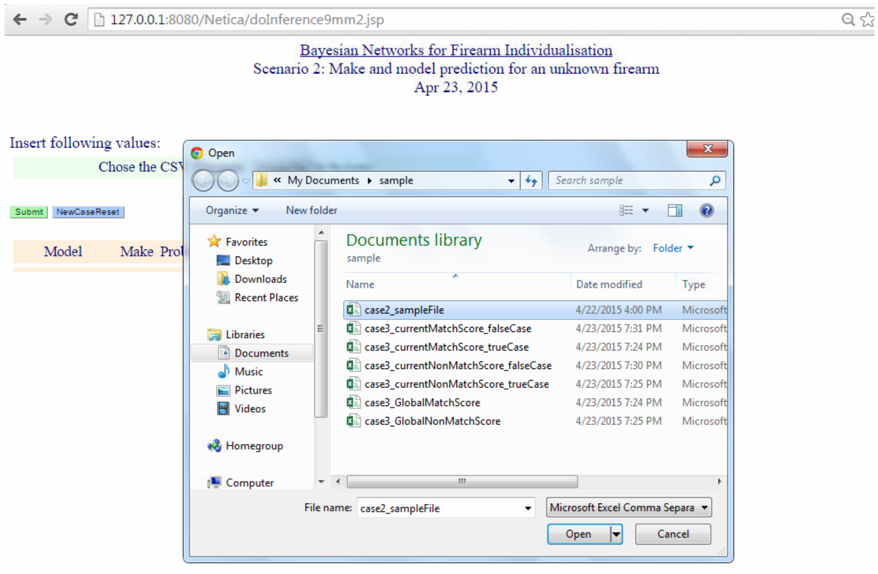

Figure 253: CSV file selection for case 2 processing via dialog box.

A dialog box should open to select the desired *.csv file. The *.csv file will contain the BF, FP, BF Rank, and FP Rank scores of the evidence cartridge case compared to the database cartridge cases. The database cartridge cases have been fired from firearms of known make and model. Once the file is located and selected, click on "Open" button (Figure 253). When the web interface is back on the screen, proceed by clicking on the green "Submit" button.

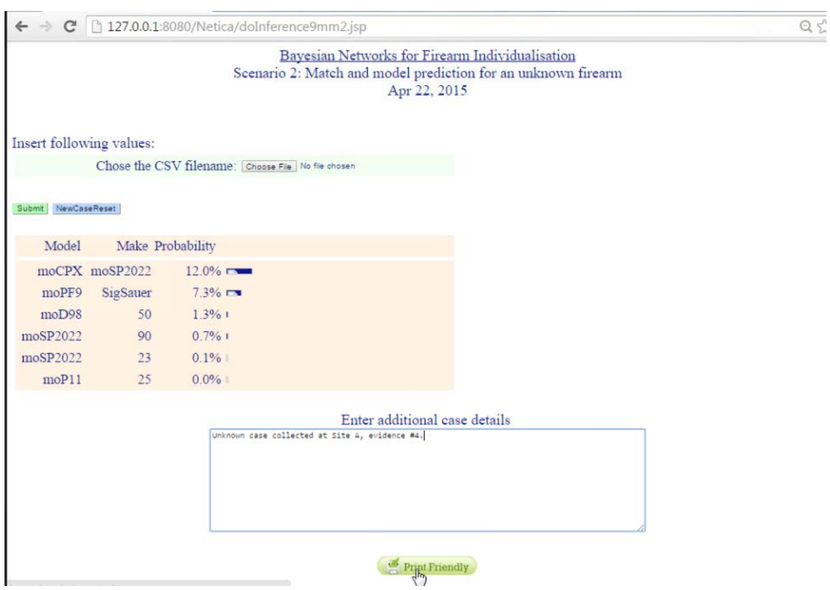

Figure 254: Ranking of the possible matches of make and model of an unknown firearm.

Clicking "Submit" will show a list, ordered by rank of make and model of firearms along with their match probability in respect to the unknown firearm from which the evidence cartridge case is likely fired (Figure 254). The best match probability was determined to be a CPX with a match probability of 12.0%. The second best result was a PF9 with a match probability of 7.3%. The examiner can input case-specific details in the provided text box. A printout of the output analysis can be done by clicking "Print Friendly."

The goal of Case 3 is to determine the likelihood ratio of a known firearm. The "Case 3 Interface" should be accessible at http://127.0.0.1:8080/Netica/doInterface9mm3.jsp.

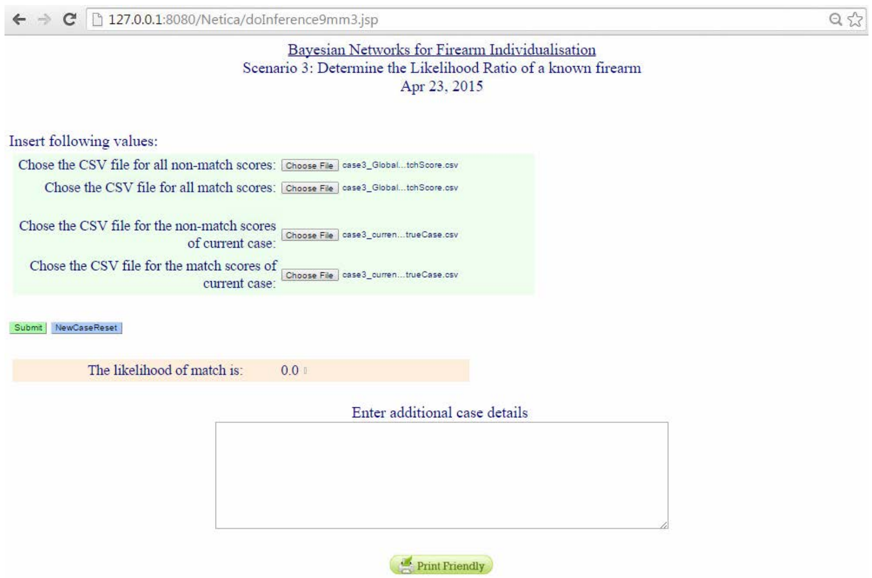

Figure 255: Case 3 web interface display after link is first accessed.

Once the web page is opened and appears as given in Figure 255, four *.csv filenames must be chosen to insert the values for the calculations. Click on "Choose File" button and upload the *.csv files, respectively to the description next to the button. Once the four desired files have been chosen, click "Submit" to upload the match score files.

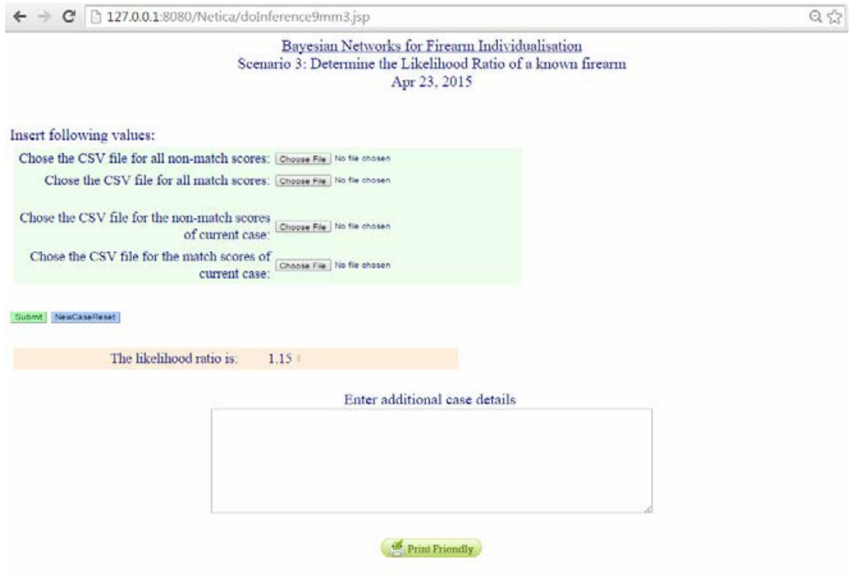

Figure 256: Calculated likelihood ratio from four match score CSV files for a known firearm.

In order to determine the likelihood ratio, each of the match score files will fit to a Gaussian mixture model. A likelihood to which the current case match scores distribution will then be estimated. A likelihood ratio will then be provided, as seen in Figure 256 as 1.15. According to Table 48, this score represents a weak to limited support of a match on the verbal likelihood scale. Finally, as in any scenario, there is the option for the examiner to add case-specific details to the designated text box and then print the output analysis by clicking “Print Friendly.”

Conclusion

A website was designed to allow for the access of the Bayesian networks by interested users.

Summary of Project Conclusions

The results of this project may be summarized as follows:

- Comparison of successively fired cartridge cases suggests, from IBIS data, that the variability between shot separations is minimal. This is probably driven by the fact that the variability within shot separations is relatively large.
- In order to perform comparisons, a firearms examiner needs to produce a certain number of test fires for purposes of comparison against an unknown cartridge case (the actual number of test fires is guided through unit policies). This research examined the question of how many cartridge cases would be representative of the firearm given the observed variability in the IBIS scores. A simulation study was performed to compare the score distributions of a randomly selected sample set (*i.e.* a set of “test fires”) against the distribution of a large sample or “estimated population” (generally 100 cartridge cases) of

a firearm. These two distributions were compared and their similarity was measured. The larger set of “test fires,” the closer the distribution of scores to that of the “population” distribution. These data suggested that the smallest sample size of test fires that would be representative of the firearm could be determined. This topic area should be researched further.

- The breech face (BF) and firing pin (FP) scores, as well as their product (BFxFP), generated by the IBIS were used to assess the ability of the system to classify an “unknown” cartridge case into a same-gun or different-gun category. There were 38 9mm Luger firearms (represented by 10 manufacturers and 18 models) used in this study. For the Ruger SR9, both the FP score and the BFxFP score were perfect classifiers. The BF score was the best classifier for four models (Glock 19, HiPoint 995TS, SigSauer SP2022, and the Taurus 24/7 G2), the FP score was the best classifier for five models (HiPoint C9, Keltec Sub 2000, Ruger SR9, SCCY CPX II, and the Springfield XD9), and the BFxFP score was the best classifier for nine models (Arcus D98, Keltec P11, Keltec PF9, Ruger LC9, Ruger P95, Ruger SR9, SigSauer P250, S&W SD9-VE, Taurus 905, and the Taurus Millennium Pro 111). The IBIS system does not provide for an easy means to use the combination of the BF and FP scores. The ability to order candidate lists through the combination of scores will be of value to firearms examiners (especially so in the 3D system). Since the markings that appear on the breech face and firing pins (or strikers) are made through independent manufacturing operations, the score generated through the IBIS comparisons are also independent. Generally, all of the classifiers performed well but the SCCY CPX II pistols were the worst in all three measures. This was due to markings that were difficult for IBIS to interpret, but would be easy identifications for a firearms examiner.
- At the request of the program manager, the IBIS was tested under both repeatability and reproducibility conditions by a number of IBIS technicians. The standard Eurachem definitions were used to assess each condition by means of the coefficient of variation (CoV). For repeatability the maximum CoV (BF) was 9% and the maximum CoV (FP) was 28%. This variability between examiners may seem high for the FP CoV, but it must be remembered that the score values obtained in this study are extremely high scores, not usually seen in casework since the same cartridge case was used in each instance. For reproducibility the maximum CoV (BF) was 11% and the maximum CoV (FP) was 29%, being very similar to the repeatability results.
- A preliminary Bayesian network was developed to assess the viability of determining the make and/or model of a firearm from the IBIS data. The results using the lowest rank, highest BF score, highest FP score, and highest BFxFP score achieved were of no significance. No further effort was expended in this direction.
- A preliminary evaluation of three blind sets was carried out at the request of the program manager. In three of the nine results an incorrect attribution was made. In one case a LR of 0.6 was obtained when the ground truth was the same gun (false negative). In another set, the ground truth of a different gun was attributed with LRs of 1.1 and 1.4 (false

positive). In retrospect, these LRs are all very close to unity, which implies that the evidence is neutral.

- In an attempt to assess the reliability of the IBIS results an expanded study of the NIST Standard Reference Material® 2461 was undertaken. Five of the NIST standards were tested under the same conditions as the reproducibility study. The IBIS was able to classify perfectly based on the BF score and the BFxFP score and almost perfectly on the FP score. Interestingly, the BF and FP scores between and within the standards ranged from 100 to 600.
- After a discussion with Ultra Electronics Forensic Technology Inc. (the producers of IBIS), a score normalization study was undertaken. Additional derived classifiers were introduced, such as FP rank, BF rank, BFxFP rank, BFxFP/BFxFP rank, normalized BF, normalized FP, and normalized BFxFP. As a result of the normalization (at a rate of 10%) there was a small improvement in the raw to normalized area under the receiver operating characteristic curve of 1.76% for BF, 2.07% for FP, and 2.16% for BFxFP. It was also found, using a Remington R1 .45 ACP pistol as an example, that generally the equal error rate improved over the sequence raw score, normalized score, and then rank. In this instance, the order of discrimination was BFFP > BF > FP. Overall it was found that a sampling rate, the proportion of the different-gun score used to determine the mean and standard deviation for the normalization process, of 20% provided the best overall results. Implementation of the normalized system for unknowns proved difficult to implement since the ground truth was unknown and the normalization depends upon knowing which of the candidates represent actual different-source firearms.
- Machine learning of the 9mm data was undertaken using techniques such as naïve Bayes, decision trees, bagged decision trees, neural networks, generalized linear model, discriminant analysis, and k-nearest neighbors. Non-match (different gun) results averaged about 98% whilst match (same-gun) averaged about 54%.
- A validation study of the proposed Bayesian network was undertaken by subdividing the data into test and training sets using random selection of samples. The test sets were run and evaluated by the ability of the network to correctly classify the sample. The averages of the areas under the curve were about 91% with a standard deviation of less than 0.2%, which decreased as the sample size increased.
- A comparison of the 2D Heritage IBIS (upon which this research is based) against that of the new 3D IBIS system (courtesy of Ultra Electronics Forensic Technology Inc., Montreal, Canada) was performed. A selection of twelve 9mm Luger firearms (representing a range of performance characteristics based on IBIS results) was used to produce a set of test cartridge cases. These cartridge cases were run through both systems. The 3D system has a number of advantages, most particularly the ability to search the side lit images. Collection of images is more time consuming (±10 minutes) as opposed to the heritage system (±3 minutes). The co-axially illuminated breech face and firing pin images yield similar results in their match scores.

- A test set was received from USACIL and analyzed using the developed Bayesian Networks. An assessment of the data is provided. After discussions with the program manager, the make and model of each firearm was provided. The test cartridge cases from the SigSauer pistols did not leave drag marks, but were of the recoil action type. The Bayesian network was updated to reflect this information. The drag mark node was split into two, viz. Drag_Mark_Sample(DB) and ActionLB_Sample(DB). Thus, from a prior odds perspective, the sample would (or not) have a drag mark, but could be from a recoil action pistol.
- Test and "evidence" samples from another DFSC study ("Baldwin study") were provided. These represented sample sets examined by firearms examiner in a "black box" type study. The data were handled in two situations. Since there were approximately 200 cartridge cases of each firearm used in the Baldwin study in the database, the comparisons were run with these data both excluded and included.
  - Excluded background: The ground truth data and the firearms examiner test results were provided. For all of the cases where examiners made an inconclusive determination the truth was that the cartridge case was fired from the same gun as the test fires (same-gun). For the examiner, false negatives Set 02 and Set 20 agreed with the examiner results. For Set 01, the correct result was achieved with LLR (Test 7)[66] being better than LLR (Test 4)[67]. In Sets 05, 09, 11, 14, 15, 17, and 18 the firearms examiners made false positive attributions. Out of the 20 comparisons, there were eight true positives, seven true negatives, five false negatives and zero false positives. In all instances of eliminations, the support for the different-gun hypothesis was, at minimum, strong.
  - When the full dataset was used the LLR (Test 7) had difficulty with Sets 01, 04, 06, and 10. For the eliminations, another candidate firearm (from the 25 firearms) tested was identified as the source of the unknown cartridge case. For Set 11, the evidence cartridge case was identified as being from a firearm outside of the original test set but the test cartridge cases were fired by a pistol within the test set.
- A website was designed to allow for the access of the Bayesian networks by interested users.

[66] LLR calculated using the firing pin type (circular or Glock-type) of the sample (unknown) cartridge case, presence of a drag mark on the sample cartridge case, the BF score, the FP score, the FP rank, and the BF rank.
[67] LLR calculated using the firing pin type (circular or Glock-type) of the sample (unknown) cartridge case, presence of a drag mark on the sample cartridge case, the BF score, and the FP score.

Appendices

Table of Figures

Table of Tables

Table of Equations

Appendix A

The code given below was used to process the IBIS reports (in *.txt format) into a *.csv file containing all relevant data (categorical and numeric). Notes and comment are given in blue, section names in red, and code in black.

```
# The purpose of this script is to process correlation reports generated by IBIS to allow for
# clean-up and conversion to .csv files for further processing in R or Netica.

mainDir <- "Z:/9mm"

# A data file is created which contains the files names generated on IBIS with the print out e.g:
# GunFile   AmmoFile    DateFile  SeqFile
# CCN        UK-SFG     313       101
# CCN       UK-SFG      313       102

# this file is then read into the data frame "info"
info<-read.csv("Z:/DataFiles/CCN.csv")

# Pay attention to the DateFile field in "info"
# Change wider to how every many characters the date contains
# For file RUG9 wider = 6

# The data from "info" are formatted into the actual file names and saved in a vector
"FileStrings"
# wider is a format size for some of the character strings in the file name

wider<-4
FileStrings <- paste(info$GunFile, info$AmmoFile, formatC(info$DateFile, width=wider, flag="0"),
formatC(info$SeqFile, width=4, flag="0"),  sep="-")

# Output directory names and paths are created
```

```
GunFile <- "CCN"
subDir <- GunFile
outDir <- "Z:/9mm File Cleanup"

# This loop will move through all of the files for a particular gun
for (t in FileStrings){

# Test code
# t<-"SCCY-PP9-101912-0145"

# the data from the IBIS output file is read in using loop t in line 29
  fname <-paste(mainDir, "/", subDir, "/", t, ".txt",sep="")
  m <- readLines(fname)

  # Test code
  # m <- readLines("Z://9mm Text Files/AR98/AR98-BZ9-101912-0060.txt")

  # the CaseID_Sample value is extracted from the file
  CaseID_Sample_str <- substr(m[2], 25, 25+20-1)

  # this section removes formatting from the file as set-up in the IBIS print out.  the grep
command
  # finds the position of each of the strings in the file.
  strings <- c("Pages", "Reference", "Case", "Information", "Reference",  "Exhibit", "Information", "Case",
"ID:", "Exhibit", "Number", "Site", "Name:", "Event:", "(Unknown)", "Law", "Agency:",
"(Unknown LAW Agency)", "Caliber:", "Acq.", "Person:", "EXAMINER",  "Comment:", "Sample", "Size",
"Tests", "ordered", "by", "Firing", "Pin", "Rank", "Breech", "Firing", "Face")

  # cut is a variable which holds information regarding the position of the character strings
  # in the strings vector which are found in the file.  Each word in strings (e.g."Case") is
  # found in the file using the grep function and stored in a temporary vector which is then
combined
  # into the vector cut.

  cut <- 0
  for (i in 1:length(strings)) {
    temp <- grep(strings[i], m)
    cut <- append(cut, temp)
  }

  # singles will contain the unique positions in the cut variable (eliminate potential repetitions)
  singles <- unique(cut)
  # the file is assigned to a new variable, k
  k <- m
```

```
  # The file k is searched through the loop, i, and all of the sub-strings will be replaced with
"NA"
   for (i in 1:length(singles)) {
      k[singles[i]] <- "NA"
   }



  # empty fields are replaced with "NA"
  k[k==""] <- "NA"

  # The final vector is created by taking everything out of k which is not "NA"
  final <- subset(k, k!="NA")

  # Clear out problem cases
  # In this section cases and exhibits entered with non-alphanumeric characters are
  # corrected to avoid issues in later processing.
  cat("Clear out problem cases","\n")
  final<-gsub("CC EX\\. ", "CCEX", final)
  final<-gsub("CC-EX\\. ", "CCEX", final)
  final<-gsub("CC EX ", "CCEX", final)
  final<-gsub("CC EX", "CCEX", final)
  final<-gsub("CC-EX ", "CC-EX", final)
  final<-gsub("CC ", "CC", final)
  final<-gsub("CC EX", "CCEX", final)
  final<-gsub("CCEX ", "CCEX", final)
  final<-gsub("CC-EXTEST 02", "CCEXTEST02", final)
  final<-gsub("GLIEE 7\\/05", "GLIEE705", final)
  final<-gsub("ITEM A", "ITEMA", final)
  final<-gsub("ITEM B", "ITEMB", final)
  final<-gsub("FIS4021111 - CW", "FIS4021111CW", final)
  final<-gsub("FIS 402 1112", "FIS4021112", final)
  final<-gsub("CASE A", "CASEA", final)
  final<-gsub("TEST 41", "TEST41", final)
  final<-gsub("TEST 45", "TEST45", final)
  final<-gsub("KTC-SUB", "SUB", final)
  final<-gsub("%", "", final)
  final<-gsub("\\/", "", final)
  final<-gsub("CCTEST 02", "CCTESTTWO", final)
  final<-gsub("NA", "0", final)

  # the trim function is defined to remove various special characters
  trim <- function (x) gsub("^\\s+|\\s+$", "", x)

  xfinal<-trim(final)

  # Multiple spaces are replaced by single spaces
  gfinal<-gsub("      ", " ", xfinal)
```

```
gfinal<-gsub("     ", " ", gfinal)
gfinal<-gsub("    ", " ", gfinal)
gfinal<-gsub("   ", " ", gfinal)
gfinal<-gsub("  ", " ", gfinal)
gfinal<-gsub(" ", ",", gfinal)

# The cleaned file is written temporarily to the drive and then read back in as a .csv file
tname<-paste(outDir, "/", subDir, "/temp.csv",sep="")
write.table(gfinal, tname,  row.names = FALSE, quote=FALSE, col.names=FALSE, eol = "\n")
gfinal<-read.csv(tname, header=FALSE)
file.remove(tname)

# The files is given column names to identify the data
colnames(gfinal) <- c("Rank", "CaseID_DB", "ExhibitNumber_DB", "SiteName", "BF", "FP", "Ejector")

# The CaseID_Sample is added to the file
CaseID_Sample<-c()

for(i in 1:length(gfinal$Rank)){
  CaseID_Sample[i]<- CaseID_Sample_str
 }

gfinal<-cbind(gfinal, CaseID_Sample)

# Reprocess File======================================
cat("Reprocess File",t,"\n")

# the basefile is created by excluding the SiteName and Ejector variable (neither are relevant
in this
# study)
basefile <- subset(gfinal,  select = c("Rank", "CaseID_DB", "ExhibitNumber_DB", "BF", "FP",
"CaseID_Sample"))

  #Case ID's and Exhibit Number==========================
cat("Case ID's and Exhibit Number","\n")

# The ExhibitNumber_Sample is the last 4 characters of the CaseID_Sample.
# These are extracted and then added as a new column
ExhibitNumber_Sample <- c()

for (i in 1:length(basefile$Rank)){
  CIDs<-as.numeric(substr(basefile$CaseID_Sample[i],17,20))
  ExhibitNumber_Sample[i] <- CIDs
}

basefile<-cbind(basefile, ExhibitNumber_Sample)
```

```
 # Case_pre_Sample and Case_pre_DB are the first 3 characters of the CaseID_Sample and
 # CaseID_DB respectively.  Identifying information about the firearms used are known and
will be
 # added to the data frame
 Case_pre_Sample <- substr(basefile$CaseID_Sample, 1, 3)
 Case_pre_DB <- substr(basefile$CaseID_DB, 1, 3)

 basefile <- cbind(basefile, Case_pre_Sample, Case_pre_DB)

 #IdentifierGun========================
 cat("IdentifierGun", "\n")

 # In Case_Sorted and Gun_ID there is a positional relationship between the two vectors.
 # Case_Sorted   Gun_ID
 # AAN          XXX724
 # AR9          XXX724
 # CAN          X66727

 Case_Sorted <- c("AAN", "AR9", "CAN", "CBN", "CCN", "CEN", "CFN", "CGN", "CHN", "CKN", "CPN",
"CQN", "CWN", "CXN", "CYN", "CZN", "FAN", "FDN", "FEN", "FJN", "FMN", "FVN", "FXN", "GAN",
"GBN", "GGN", "GNN", "GSN", "GWN", "GXN", "HBN", "HCN", "HFN", "HIC", "HIP", "HJN", "HKN",
"HSN", "HTN", "HVN", "HWN", "HXN", "KTC", "RAN", "RBN", "RCN", "RFN", "RNN", "RPN", "RUG",
"RVN", "SCC", "SFX", "SIG", "SUB", "SWS", "TAN", "TGN", "TNN", "TPN", "TRM", "TRT", "TTN",
"TVN", "TWN", "TXN", "UPN", "WAN", "WBN", "WCN", "WDN", "WEN", "WFN", "WPN", "WSN", "WVN",
"WWN", "WXN")

 # The gun_ID are the last 5 characters of the serial number of a particular firearm.  All
leading zero are
 # replaced with an "X".  An "X" is also prefixed to all of the Gun_ID's.  Each of these
identifiers is
 # unique.
 Gun_ID <- c("XXX724", "XXX724", "X66727", "X97569", "X66727", "X97570", "X97568", "X97571",
"X97569", "X66727", "X66727", "X97568", "X97570", "X66727", "X97571", "X66727", "X17849",
"X77862", "X17802", "X77862", "X77862", "X17841", "X77862", "XLB713", "XTE408", "XLB713",
"XLB713", "XLB713", "XAS648", "XAS012", "XX9554", "X80728", "XX9554", "XX9554", "X80728",
"X80728", "X55429", "XX9554", "XX9554", "X55420", "X55457", "X55426", "XSBP59", "X43521",
"X32446", "X33654", "X44279", "X44279", "X44279", "X69363", "X44279", "X66727", "X77862",
"X82066", "XEF603", "XA9892", "X45401", "X55720", "X20246", "X55720", "X54042", "X55720",
"X55720", "X45399", "X45405", "X45398", "XX5056", "XEF603", "XSHQ79", "XSJN79", "XSJP08",
"XEF603", "XSBP59", "XEF603", "XEF603", "XAZV54", "XAZV54", "XSHQ08")

  # The IdentifierGun_Sample and the IdentifierGun_DB vectors are initialized
 IdentifierGun_Sample <- c()
 IdentifierGun_DB <- c()
```

```
 # In this loop the file will be evaluated at each row for the value of Case_pre_Sample (see
line 167)
 # This takes place while looping through all of the values of the Case_Sorted vector.  If the
 # Case_pre_Sample value is the same as the Case_Sorted value then they are from the same
firearm.
 # The index of the value in Case_Sorted applies to the same index value in Gun_ID.  The i-th
value of
 # IdentifierGun_Sample is thus assigned the j-th value of Gun_ID (see line 214).  A similar
process
 # occurs for IdentifierGun_DB.  In the instance where the information for the database is
unknown,
 # the following occurs: In line 213, a flag (label_DB) is set to zero.  Whenever an assignment
is
 # made to IdentifierGun_DB, then flag is changed to one (line 220).  After the comparisons
are
 # completed a test is made for the value of the flag (line 226).  If this test is true (label_DB <
1),
 # then the value "Unknown" is assigned to the i-th value of IdentifierGun_DB
 for (i in 1:length(basefile$Rank)){

   label_DB <- 0

   for (j in 1:length(Gun_ID)){
     if(Case_pre_Sample[i] == Case_Sorted[j]){IdentifierGun_Sample[i] <-Gun_ID[j]}
     if(Case_pre_DB[i] == Case_Sorted[j]){
       IdentifierGun_DB[i] <- Gun_ID[j]
       label_DB <- 1
     }
   }
   if(label_DB<1){IdentifierGun_DB[i] <- "Unknown"}
 }

 basefile<-cbind(basefile, IdentifierGun_Sample, IdentifierGun_DB)

 #25 SR9's===============================
 # This process adds the identifiers for the 25 SR9's used in the Baldwin Study.
 # It is performed differently since the file naming sequence for these files is slightly different.

 cat("SR9's","\n")
 Identifier <- c("RN-JK01", "RN-JK02", "RN-JK03", "RN-JK04", "RN-JK05", "RN-JK06", "RN-JK07",
 "RN-JK08", "RN-JK09", "RN-JK10", "RN-JK11", "RN-JK12", "RN-JK13", "RN-JK14", "RN-JK15",
 "RN-JK16", "RN-JK17", "RN-JK18", "RN-JK19", "RN-JK20", "RN-JK21", "RN-JK22", "RN-JK23",
 "RN-JK24", "RN-JK25")
```

```
  IdentifierGun <- c("X96383", "X96385", "X96387", "X96388", "X96584", "X96585", "X96586",
"X96590",
  "X96592", "X96593", "X96594", "X96604", "X96620", "X96649", "X96651", "X96661", "X96663",
  "X96664", "X96665", "X96667", "X96669", "X96681", "X96689", "X96718", "X96719")

  # In this instance, new values will be added to the IdentifierGun_Sample and
IdentifierGun_DB columns
  # of the basefile data frame.  In order to be added, these values need to be allowed (form part
of the
  # levels of that column).  The existing levels are extracted (line 242), and the new levels are
added (line
  # 243)
  old_levels<-levels(basefile$IdentifierGun_Sample)
  levels(basefile$IdentifierGun_Sample) <- c(old_levels,IdentifierGun)

  old_levels<-levels(basefile$IdentifierGun_DB)
  levels(basefile$IdentifierGun_DB) <- c(old_levels,IdentifierGun)

  # the sample and db values are equivalent in structure to the values in Identifier (line 236) to
allow for
  # comparison.
  sample <-substr(basefile$CaseID_Sample,1, 7)
  db <-substr(basefile$CaseID_DB,1, 7)

  counter<-length(Identifier)

  for (i in 1:counter){

    out_sample<-grep(Identifier[i], sample)
    basefile$IdentifierGun_Sample[out_sample]<-IdentifierGun[i]

    out_db<-grep(Identifier[i], db)
    basefile$IdentifierGun_DB[out_db]<-IdentifierGun[i]

  }



  #Match=================================
  cat("Match", "\n")

  # Matches are easily determined. If the i-th value of IdentifierGun_Sample and the i-th value
of
  # IdentifierGun_DB are the same, then Match=Yes (same gun).
  Match<-c()

  for (i in 1:length(basefile$Rank)){
```

```
      if(as.character(basefile$IdentifierGun_Sample[i])==
       as.character(basefile$IdentifierGun_DB[i])){Match[i]<-"Yes"}else{Match[i]<-"No"}
 }

 basefile<-cbind(basefile,Match)

 #Makes & Models================================
 cat("Makes & Models", "\n")

 # A similar process is used as described above for the Makes and Models of the firearms.
 # There is a unique directional relationship from Identifier to Model to Make.
 Make<-c("SigSauer", "Springfield", "Springfield", "Springfield", "Taurus", "Ruger", "Ruger", "Ruger",
"Ruger", "Taurus", "Taurus", "Taurus", "Taurus", "Taurus", "Hi-Point", "Hi-Point", "Hi-Point", "Hi-Point",
"Taurus", "SCCY", "Ruger", "Springfield", "Hi-Point", "SigSauer", "SCCY", "SCCY", "SCCY", "SCCY",
"Smith&Wesson", "Glock", "Glock", "Keltec", "Keltec", "Glock", "Keltec", "Keltec", "Keltec", "Keltec",
"Keltec", "Glock", "Hi-Point", "Arcus")

 Model<-c("SP2022", "XD9", "XD9", "XD9", "905",  "LC9", "LC9", "LC9", "P95", "24/7G2", "24/7G2",
"24/7G2", "24/7G2", "MillenniumPro111", "C9", "C9", "C9", "C9", "24/7G2", "CPX", "SR9", "XD9", "C9",
"P250", "CPX", "CPX", "CPX", "CPX", "SD9-VE","19Gen4","19Gen4","P-11","Sub-2000", "19Gen4",
"PF9", "PF9", "PF9", "PF9", "PF9", "19Gen4", "995TS", "D98")

Identifier<- c("X05056", "X17802", "X17841", "X17849", "X20246", "X32446", "X33654", "X43521",
"X44279", "X45398", "X45399", "X45401", "X45405", "X54042", "X55420", "X55426", "X55429",
"X55457", "X55720", "X66727", "X69363", "X77862", "X80728", "X82066", "X97568", "X97569",
"X97570", "X97571", "XA9892", "XAS012", "XAS648", "XAZV54", "XEF603", "XLB713", "XSBP59",
"XSHQ08", "XSHQ79", "XSJN79", "XSJP08", "XTE408",  "XX9554", "XXX724")

 Model_Sample <- c()
 Make_Sample <- c()
 Model_DB <- c()
 Make_DB <- c()

 for (i in 1:length(basefile$Rank)){

   label_DB<-0

     for (j in 1:length(Identifier)){
     if(basefile$IdentifierGun_Sample[i] == Identifier[j]){
       Model_Sample[i] <- Model[j]
       Make_Sample[i] <- Make[j]
     }

     if(basefile$IdentifierGun_DB[i] == Identifier[j]){
       Model_DB[i] <- Model[j]
       Make_DB[i] <- Make[j]

       label_DB<-1
     }
```

```
    }
    if(label_DB<1){
      Model_DB[i] <- "Unknown"
      Make_DB[i] <- "Unknown"}
  }

  basefile <- cbind(basefile, Model_Sample, Make_Sample, Model_DB, Make_DB)

  # Drag Marks and Firearm Type============================
  cat("Drag Marks and Firearm Type", "\n")

  # A similar process is used as described above for the Drag Marks and Firearm Types of the
firearms.
  # There is a unique directional relationship from Identifier to Type, and Identifier to Drag.

Identifier <- c("AAN", "AR9", "CAN", "CBN", "CCN", "CEN", "CFN", "CGN", "CHN", "CKN", "CPN",
"CQN", "CWN", "CXN", "CYN", "CZN", "FAN", "FDN", "FEN", "FJN", "FMN", "FVN", "FXN", "GAN",
"GBN", "GGN", "GNN", "GSN", "GWN", "GXN", "HBN", "HCN", "HFN", "HIC", "HIP", "HJN", "HKN",
"HSN", "HTN", "HVN", "HWN", "HXN", "KTC", "RAN", "RBN", "RCN", "RFN", "RNN", "RPN", "RUG",
"RVN", "SCC", "SFX", "SIG", "SUB", "SWS", "TAN", "TGN", "TNN", "TPN", "TRM", "TRT", "TTN",
"TVN", "TWN", "TXN", "UPN", "WAN", "WBN", "WCN", "WDN", "WEN", "WFN", "WPN", "WSN", "WVN",
"WWN", "WXN", "TRT", "RN-")

 Type <- c("Pistol", "Pistol", "Pistol", "Pistol", "Pistol", "Pistol", "Pistol", "Pistol", "Pistol", "Pistol", "Pistol",
"Pistol", "Pistol", "Pistol", "Pistol", "Pistol", "Pistol", "Pistol", "Pistol", "Pistol", "Pistol", "Pistol", "Pistol",
"Pistol", "Pistol", "Pistol", "Pistol", "Pistol", "Pistol", "Pistol", "Carbine", "Pistol", "Carbine", "Carbine",
"Pistol", "Pistol", "Pistol", "Carbine", "Carbine", "Pistol", "Pistol", "Pistol", "Pistol", "Pistol", "Pistol",
"Pistol", "Pistol", "Pistol", "Pistol", "Pistol", "Pistol", "Pistol", "Pistol", "Pistol", "Carbine", "Pistol", "Pistol",
"Pistol", "Revolver", "Pistol", "Pistol", "Pistol", "Pistol", "Pistol", "Pistol", "Pistol", "Pistol", "Carbine",
"Pistol", "Pistol", "Pistol", "Carbine", "Pistol", "Carbine", "Carbine", "Pistol", "Pistol", "Pistol", "Pistol",
"Pistol")

Drag <- c("Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes",
"Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes",
"Yes", "Yes", "No", "No", "No", "No", "No", "No", "No", "No", "No", "No", "No", "No", "Yes", "Yes", "Yes",
"Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "No", "Yes", "Yes", "Yes", "No", "Yes",
"Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "No", "Yes", "Yes", "Yes", "No", "Yes", "No", "No",
"Yes", "Yes", "Yes", "Yes", "Yes")

  Type_Sample <- c()
  Type_DB <-c()
  Drag_Mark_Sample <- c()
  Drag_Mark_DB <- c()

  for (i in 1:length(basefile$Rank)){

    label_DB<-0

    for (j in 1:length(Identifier)){
```

```
      if(basefile$Case_pre_Sample[i] == Identifier[j]){
        Type_Sample[i] <- Type[j]
        Drag_Mark_Sample[i] <- Drag[j]
      }
      if(basefile$Case_pre_DB[i] == Identifier[j]){
        Type_DB[i] <- Type[j]
        Drag_Mark_DB[i] <- Drag[j]

        label_DB<-1
      }
    }
    if(label_DB<1){
      Type_DB[i] <- "Unknown"
      Drag_Mark_DB[i] <- "Unknown"}
  }

  basefile <- cbind(basefile, Type_Sample, Type_DB, Drag_Mark_Sample, Drag_Mark_DB)

  # Ammo and Primers=============================
  cat("Ammo and Primers", "\n")

  # A similar process is used as described above for the Ammo and Primers used in the test
fires.
  # There is a unique directional relationship from Identifier to Ammo, and Identifier to Primer.
  # In this instance, if the Ammo is known, the primer is unknown and vice-versa.

Ammo <- c("Unknown", "Unknown", "Unknown", "Unknown", "Unknown", "Unknown", "Unknown",
"Unknown", "Unknown", "Unknown", "Unknown", "Unknown", "FederalAmericanEagle", "Unknown",
"Blazer", "FederalPremium", "FederalAmericanEagle", "Sellier&Bellot", "Lapua", "Blazer", "Unknown",
"Unknown", "Armscor", "FederalAmericanEagle", "PrviPartizan", "Winchester", "Unknown", "Unknown",
"Unknown", "Unknown", "Unknown", "Unknown", "Unknown", "Unknown", "Unknown", "Unknown",
"Unknown", "Unknown", "FederalAmericanEagle", "Unknown", "Unknown", "Blazer", "Unknown",
"Unknown", "Blazer", "Blazer", "Blazer", "Blazer", "Blazer", "Blazer", "Blazer", "Blazer", "Blazer",
"Unknown", "FederalAmericanEagle", "Blazer", "Unknown", "Unknown", "Unknown", "Unknown",
"Unknown", "Unknown", "Unknown", "Armscor", "Unknown", "Unknown", "Unknown", "Unknown",
"Unknown", "FederalAmericanEagle", "Unknown", "Unknown", "FederalAmericanEagle", "Unknown",
"Remington")

Primer <- c("SSG", "STP", "SFT", "SFG", "STP", "SSG", "SRG", "STP", "SCT", "SCT", "SCT", "SCT",
"Unknown", "SRP", "Unknown", "Unknown", "Unknown", "Unknown", "Unknown", "Unknown", "SRP",
"Unknown", "Unknown", "Unknown", "Unknown", "Unknown", "SRP", "SFT", "SFT", "SCT", "SCT",
"SCT", "SCT", "SSG", "SSG", "SFG", "SRG", "STP", "Unknown", "STP", "SFG", "Unknown", "SSG",
"SRG", "Unknown", "Unknown", "Unknown", "Unknown", "Unknown", "Unknown", "Unknown",
"Unknown", "Unknown", "SRP", "Unknown", "Unknown", "SFT", "SRT", "SRT", "SFG", "STP", "SRG",
"SSG", "Unknown", "SSG", "STP", "STP", "STP", "SCT", "Unknown", "STP", "SRG", "Unknown", "SFG",
"Unknown")

Identifier <- c("AAN", "GXN", "GWN", "GAN", "GGN", "GNN", "GSN", "GBN", "HVN", "HXN", "HKN",
```

```
"HWN", "HCN", "HJN", "HBN", "HFN", "HIC", "HSN", "HTN", "WVN", "WWN", "KTC", "WAN", "WEN",
"WPN", "WSN", "WFN", "WXN", "WBN", "WCN", "WDN", "RBN", "RCN", "RAN", "RFN", "RNN", "RPN",
"RVN", "RUG", "CAN", "CCN", "CKN", "CPN", "CXN", "CZN", "CFN", "CQN", "CBN", "CHN", "CEN",
"CWN", "CGN", "CYN", "UPN", "SIG", "SWS", "FEN", "FVN", "FAN", "FDN", "FJN", "FMN", "FXN",
"SFX", "TNN", "TXN", "TVN", "TAN", "TWN", "TRM", "TGN", "TPN", "TRT", "TTN", "RN-")


 Ammo_Sample <- c()
 Ammo_DB <-c()
 Primer_Sample <- c()
 Primer_DB <- c()

 for (i in 1:length(basefile$Rank)){

  label_DB<-0

  for (j in 1:length(Identifier)){
   if(basefile$Case_pre_Sample[i] == Identifier[j]){
    Ammo_Sample[i] <- Ammo[j]
    Primer_Sample[i] <- Primer[j]
   }
   if(basefile$Case_pre_DB[i] == Identifier[j]){
    Ammo_DB[i] <- Ammo[j]
    Primer_DB[i] <- Primer[j]

    label_DB<-1
   }
  }
  if(label_DB<1){
   Ammo_DB[i] <- "Unknown"
   Primer_DB[i] <- "Unknown"}
 }

 old<-Ammo_DB

 #Mixed Ammo============================
 # For certain initial firings mixed ammunition was used in similar strings.  This routine
accounts
 # for these firings using a similar procedure.

 cat("Mixed Ammo", "\n")

 caseString<-c("98-AC9", "C-AC9-", "CY-AC9", "N-AR-U", "PO-AC9", "XD-AC9", "98-BZ9", "C-B29-",
 "C-BZR9", "CY-BZ9", "N-BB-U", "PO-BZ9", "S9-BLZ", "XD-BZ9", "C-CBC9", "98-FA9", "CA-FC9",
 "C-FA9-",  "C-SUB-", "CY-FA9", "G9-FAE", "M9-FAE", "N-FA-U", "PO-FA9", "T9-FAE", "XD-FA9",
 "98-FC9", "C-FC9-", "C-FCD9", "CY-FC9", "PO-FC9", "XD-FC9", "C-GFL9", "98-HC9", "C-HC9-",
 "CY-HC9", "PO-HC9", "XD-HC9", "98-LP9", "CY-LP9", "N-LP-U", "PO-LP9", "XD-LP9", "C-PMC9",
 "98-PP9", "C-PPU9", "N-PP-U", "PO-PP9", "XD-PP9", "C-RP9-", "-JK01-", "-JK02-", "-JK03-", "-JK04-",
 "-JK05-", "-JK06-", "-JK07-", "-JK08-", "-JK09-", "-JK10-", "-JK11-", "-JK12-", "-JK13-", "-JK14-", "-JK15-
```

```
",
 "-JK16-", "-JK17-", "-JK18-","-JK19-","-JK20-","-JK21-","-JK22-", "-JK23-", "-JK24-", "-JK25-", "98-SB9",
 "C-SB9-", "C-SB99", "CY-SB9", "N-SB-U", "PO-SB9", "XD-SB9", "C-SP9-", "C-SP49", "98-SG9",
 "CY-SG9", "XD-SG9", "PO-SG9", "C-WC69", "C-WC79", "C-WC89", "C-WIN9", "N-WC-U", "98-WX9",
 "C-WX9-", "CY-WX9", "PO-WX9", "XD-WX9", "CY-PP9")

  Ammo<-c("Armscor", "Armscor", "Armscor", "Armscor", "Armscor", "Armscor", "Blazer", "Blazer",
"Blazer", "Blazer", "Blazer", "Blazer", "Blazer", "Blazer", "CBC", "FederalAmericanEagle",
"FederalAmericanEagle", "FederalAmericanEagle", "FederalAmericanEagle", "FederalAmericanEagle",
"FederalAmericanEagle", "FederalAmericanEagle", "FederalAmericanEagle", "FederalAmericanEagle",
"FederalAmericanEagle", "FederalAmericanEagle", "FederalChampion", "FederalChampion",
"FederalChampion", "FederalChampion", "FederalChampion", "FederalChampion", "GFL",
"HornadyCriticalDuty", "HornadyCriticalDuty", "HornadyCriticalDuty", "HornadyCriticalDuty",
"HornadyCriticalDuty", "Lapua", "Lapua", "Lapua", "Lapua", "Lapua", "PMC", "PrviPartizan",
"PrviPartizan", "PrviPartizan", "PrviPartizan", "PrviPartizan", "Remington", "Remington", "Remington",
"Remington", "Remington", "Remington", "Remington", "Remington", "Remington", "Remington",
"Remington", "Remington", "Remington", "Remington", "Remington", "Remington", "Remington",
"Remington", "Remington", "Remington", "Remington", "Remington", "Remington", "Remington",
"Remington", "Remington", "Sellier&Bellot", "Sellier&Bellot", "Sellier&Bellot", "Sellier&Bellot",
"Sellier&Bellot", "Sellier&Bellot", "Sellier&Bellot", "Speer", "Speer004", "SpeerGoldDot",
"SpeerGoldDot", "SpeerGoldDot", "SpeerGoldDot", "WCC06", "WCC07", "WCC08", "WIN",
"Winchester", "WinchesterSuperX", "WinchesterSuperX", "WinchesterSuperX", "WinchesterSuperX",
"WinchesterSuperX", "PrviPartizan")

  ammoString_Sample<-substr(basefile$CaseID_Sample,3,8)
  ammoString_DB<-substr(basefile$CaseID_DB,3,8)

  for (i in 1:length(basefile$Rank)){

    for (j in 1:length(caseString)){
      if(ammoString_Sample[i] == caseString[j]){
        Ammo_Sample[i] <- Ammo[j]
        #cat(ammoString_Sample[i],caseString[j],"Match Sample","\n")
        Primer_Sample[i] <- "Unknown"
      }

      if(ammoString_DB[i] == caseString[j]){
        Ammo_DB[i] <- Ammo[j]
        #cat(ammoString_DB[i],caseString[j],"Match DB","\n")
        Primer_DB[i] <- "Unknown"
      }
    }
  }

  basefile <- cbind(basefile, Ammo_Sample, Ammo_DB, Primer_Sample, Primer_DB)

  #Caliber=================================
  cat("Caliber", "\n")
```

```
# In this instance all of the calibers are the same.
CaliberGun <- c("9mm", "9mm", "9mm", "9mm", "9mm", "9mm", "9mm", "9mm", "9mm", "9mm",
"9mm", "9mm", "9mm", "9mm", "9mm", "9mm", "9mm", "9mm", "9mm", "9mm", "9mm", "9mm", "9mm",
"9mm", "9mm", "9mm", "9mm", "9mm", "9mm", "9mm", "9mm", "9mm", "9mm", "9mm", "9mm", "9mm",
"9mm", "9mm", "9mm", "9mm", "9mm", "9mm", "9mm", "9mm", "9mm", "9mm", "9mm", "9mm", "9mm","
9mm", "9mm", "9mm", "9mm", "9mm", "9mm", "9mm", "9mm", "9mm", "9mm", "9mm", "9mm", "9mm",
"9mm", "9mm", "9mm", "9mm", "9mm", "9mm", "9mm", "9mm", "9mm", "9mm", "9mm", "9mm", "9mm",
"9mm", "9mm", "9mm")

Identifier<-c("AAN", "AR9", "CAN"," CBN", "CCN", "CEN", "CFN", "CGN", "CHN", "CKN", "CPN",
"CQN", "CWN", "CXN", "CYN", "CZN", "FAN", "FDN", "FEN", "FJN", "FMN", "FVN", "FXN", "GAN",
"GBN", "GGN", "GNN", "GSN", "GWN", "GXN", "HBN", "HCN", "HFN", "HIC", "HIP", "HJN", "HKN",
"HSN", "HTN", "HVN", "HWN", "HXN", "KTC", "RAN", "RBN", "RCN", "RFN", "RNN", "RPN", "RUG",
"RVN", "SCC", "SUB", "TAN", "TGN", "TNN", "TPN", "TRM", "TRT", "TTN", "TVN", "TWN", "TXN",
"UPN", "WAN", "WBN", "WCN", "WDN", "WEN", "WFN", "WPN", "WSN", "WVN", "WWN", "WXN" ,"RN-
", "SIG", "SFX")

Caliber_Sample<-c()
Caliber_DB<-c()

for (i in 1:length(basefile$Rank)){

  for (j in 1:length(Identifier)){
    if(basefile$Case_pre_Sample[i] == Identifier[j]){
      Caliber_Sample[i] <- CaliberGun[j]
    }

    if(basefile$Case_pre_DB[i] == Identifier[j]){
      Caliber_DB[i] <- CaliberGun[j]
    }else{
      Caliber_DB[i] <- "9mm"
    }

  }
}

basefile <- cbind(basefile, Caliber_Sample, Caliber_DB)

#FiringPinType=================================
cat("FiringPinType", "\n")

Identifier <- c("AAN", "AR9", "GXN", "GWN", "GAN", "GGN", "GNN", "GSN", "GBN", "HVN", "HXN",
"HKN", "HWN", "HCN", "HIP", "HJN", "HBN", "HFN"," HIC", "HSN", "HTN", "WVN", "WWN", "KTC",
"WAN", "WEN", "WPN", "WSN", "KTC", "WFN", "WXN", "WBN", "WCN", "WDN", "RBN", "RCN", "RAN",
"RFN", "RNN", "RPN", "RVN", "RUG"," CAN", "CCN", "CKN", "CPN", "CXN", "CZN", "SCC", "CFN",
"CQN", "CBN", "CHN", "CEN"," CWN", "CGN", "CYN", "UPN", "SIG", "SWS", "FEN", "FVN", "FAN",
"FDN", "FJN", "FMN", "FXN", "SFX", "TNN", "TXN", "TVN", "TAN", "TWN", "TRM", "TGN", "TPN","
TRT", "TTN", "RN-")
```

```
FiringPin <- c("Circular", "Circular", "Glock", "Glock", "Glock", "Glock", "Glock", "Glock", "Glock",
"Circular", "Circular", "Circular", "Circular", "Circular", "Circular", "Circular", "Circular", "Circular",
"Circular", "Circular", "Circular", "Circular", "Circular", "Circular", "Circular", "Circular", "Circular",
"Circular", "Circular", "Circular", "Circular", "Circular", "Circular", "Circular", "Circular", "Circular",
"Circular", "Circular", "Circular", "Circular", "Circular", "Circular", "Circular", "Circular", "Circular",
"Circular", "Circular", "Circular", "Circular", "Circular", "Circular", "Circular", "Circular", "Circular",
"Circular", "Circular", "Circular", "Circular", "Circular", "Circular", "Circular", "Circular", "Circular",
"Circular", "Circular", "Circular", "Circular", "Circular", "Circular", "Circular", "Circular", "Circular",
"Circular", "Circular", "Circular", "Circular", "Circular", "Circular", "Circular")

 Firing_Pin_Type_Sample <-c()
 Firing_Pin_Type_DB<-c()

 for (i in 1:length(basefile$Rank)){

  label_DB<-0

  for (j in 1:length(Identifier)){
   if(basefile$Case_pre_Sample[i] == Identifier[j]){
    Firing_Pin_Type_Sample[i] <- FiringPin[j]

   }

   if(basefile$Case_pre_DB[i] == Identifier[j]){
    Firing_Pin_Type_DB[i] <- FiringPin[j]

    label_DB<-1
   }
   if(label_DB<1){Firing_Pin_Type_DB[i] <- "Circular"}

  }
 }

 basefile <- cbind(basefile, Firing_Pin_Type_Sample, Firing_Pin_Type_DB)

 #Reload=================================
 cat("Reload", "\n")

 Reloader <- c ("Yes", "No", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes",
"No", "No", "Yes", "No", "No", "No", "No", "No", "No", "Yes", "No", "No", "No", "No", "No", "No", "Yes",
"Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "No", "Yes", "Yes", "No",
"Yes", "Yes", "No", "No", "No", "No", "No", "No", "No", "No", "No", "No", "Yes", "No"," No", "Yes", "Yes",
"Yes", "Yes", "Yes", "Yes", "Yes", "No", "Yes", "Yes", "Yes", "Yes", "Yes", "No", "Yes", "Yes", "No",
"Yes", "No")

 Identifier <- c("AAN", "AR9", "GXN", "GWN", "GAN", "GGN", "GNN"," GSN", "GBN", "HVN", "HXN",
"HKN", "HWN", "HCN", "HIP", "HJN", "HBN", "HFN", "HIC", "HSN", "HTN", "WVN", "WWN", "KTC",
"WAN", "WEN", "WPN", "WSN", "KTC", "WFN", "WXN", "WBN", "WCN", "WDN", "RBN", "RCN", "RAN",
"RFN", "RNN", "RPN", "RVN", "RUG", "CAN", "CCN", "CKN", "CPN", "CXN", "CZN", "SCC", "CFN",
"CQN", "CBN", "CHN", "CEN", "CWN", "CGN", "CYN", "UPN", "SIG", "SWS", "FEN", "FVN", "FAN",
```

```
"FDN", "FJN", "FMN", "FXN", "SFX", "TNN", "TXN", "TVN", "TAN", "TWN", "TRM", "TGN", "TPN",
"TRT", "TTN", "RN-")

 Reload <-c()

 for (i in 1:length(basefile$Rank)){

   for (j in 1:length(Identifier)){
     if(basefile$Case_pre_Sample[i] == Identifier[j]){
       Reload[i] <- Reloader[j]

     }
   }
 }

 basefile <- cbind(basefile, Reload)

 # Clear out problem cases=================================
 # In these files there are particular entries which are problematic.
 # Some information about the firearms in the database is known, but it is incomplete.
 # As discussed in line 239, new levels may need to be added.
 cat("Clear out problem cases", "\n")

 old_levels<-levels(basefile$Model_DB)
 levels(basefile$Model_DB) <- c(old_levels,"P85","SR9","P95")

 old_levels<-levels(basefile$Make_DB)
 levels(basefile$Make_DB) <- c(old_levels,"Ruger")

 old_levels<-levels(basefile$Drag_Mark_DB)
 levels(basefile$Drag_Mark_DB) <- c(old_levels,"Yes")

 out<-0
 out<-grep("ELS(.*)668", basefile$CaseID_DB)
 basefile$Model_DB[out]<-"SR9"
 basefile$Make_DB[out]<-"Ruger"
 basefile$Drag_Mark_DB[out]<-"Yes"

 out<-0
 out<-grep("ELS(.*)880", basefile$CaseID_DB)
 basefile$Model_DB[out]<-"SR9"
 basefile$Make_DB[out]<-"Ruger"
 basefile$Drag_Mark_DB[out]<-"Yes"

 out<-0
 out<-grep("ELS(.*)693", basefile$CaseID_DB)
 basefile$Model_DB[out]<-"P85"
 basefile$Make_DB[out]<-"Ruger"
 basefile$Drag_Mark_DB[out]<-"Yes"
```

```
out<-0
out<-grep("DW(.*)693", basefile$CaseID_DB)
basefile$Model_DB[out]<-"P85"
basefile$Make_DB[out]<-"Ruger"
basefile$Drag_Mark_DB[out]<-"Yes"

out<-0
out<-grep("KEE(.*)693", basefile$CaseID_DB)
basefile$Model_DB[out]<-"P85"
basefile$Make_DB[out]<-"Ruger"
basefile$Drag_Mark_DB[out]<-"Yes"

out<-0
out<-grep("WID(.*)693", basefile$CaseID_DB)
basefile$Model_DB[out]<-"P85"
basefile$Make_DB[out]<-"Ruger"
basefile$Drag_Mark_DB[out]<-"Yes"

out<-0
out<-grep("ELS(.*)88", basefile$CaseID_DB)
basefile$Model_DB[out]<-"SR9"
basefile$Make_DB[out]<-"Ruger"
basefile$Drag_Mark_DB[out]<-"Yes"

out<-0
out<-grep("KEE(.*)88", basefile$CaseID_DB)
basefile$Model_DB[out]<-"SR9"
basefile$Make_DB[out]<-"Ruger"
basefile$Drag_Mark_DB[out]<-"Yes"

out<-0
out<-grep("POW(.*)88", basefile$CaseID_DB)
basefile$Model_DB[out]<-"SR9"
basefile$Make_DB[out]<-"Ruger"
basefile$Drag_Mark_DB[out]<-"Yes"

out<-0
out<-grep("WID(.*)88", basefile$CaseID_DB)
basefile$Model_DB[out]<-"SR9"
basefile$Make_DB[out]<-"Ruger"
basefile$Drag_Mark_DB[out]<-"Yes"

out<-0
out<-grep("DW(.*)668", basefile$CaseID_DB)
basefile$Model_DB[out]<-"SR9"
basefile$Make_DB[out]<-"Ruger"
basefile$Drag_Mark_DB[out]<-"Yes"

out<-0
```

```
  out<-grep("ELS(.*)668", basefile$CaseID_DB)
  basefile$Model_DB[out]<-"SR9"
  basefile$Make_DB[out]<-"Ruger"
  basefile$Drag_Mark_DB[out]<-"Yes"

  out<-0
  out<-grep("DW(.*)668", basefile$CaseID_DB)
  basefile$Model_DB[out]<-"SR9"
  basefile$Make_DB[out]<-"Ruger"
  basefile$Drag_Mark_DB[out]<-"Yes"

  out<-0
  out<-grep("P85", basefile$CaseID_DB)
  basefile$Model_DB[out]<-"P85"
  basefile$Make_DB[out]<-"Ruger"
  basefile$Drag_Mark_DB[out]<-"Yes"

  out<-0
  out<-grep("P95", basefile$CaseID_DB)
  basefile$Model_DB[out]<-"P95"
  basefile$Make_DB[out]<-"Ruger"
  basefile$Drag_Mark_DB[out]<-"Yes"

  out<-0
  out<-grep("RN-JK", basefile$CaseID_DB)
  basefile$Model_DB[out]<-"SR9"
  basefile$Make_DB[out]<-"Ruger"
  basefile$Drag_Mark_DB[out]<-"Yes"

  #Write File=======================
  cat("Write File", "\n")

  # The data frame is re-ordered and the data required in the final file is included.
  # The updated data frame is then written to a .csv file.

out <-subset(basefile, select = c(CaseID_Sample, ExhibitNumber_Sample, Rank, CaseID_DB,
ExhibitNumber_DB, BF, FP, Match, Make_DB, Model_DB, Ammo_DB, Caliber_DB,
Firing_Pin_Type_DB, Make_Sample, Model_Sample, Ammo_Sample, IdentifierGun_Sample,
IdentifierGun_DB, Caliber_Sample, Firing_Pin_Type_Sample, Type_Sample, Primer_Sample,
Primer_DB, Drag_Mark_Sample, Drag_Mark_DB, Reload))

  zname<-paste(outDir,"/", subDir,"/",t,"Clean.csv",sep="")

  write.csv(out, zname, row.names = FALSE)

}
```

Appendix B

Script used to create a Bayesian network for Netica® through RNetica.

```
NeticaLicenseKey <- "insert valid license key"
library(RNetica)

# Load data
==============================================================================
Firearms <- read.csv("Z:/CompleteNineMM_July2015DragUpdated.csv")

# Discretize FP, BF, Rank and Rank_BF
=======================================================
FirearmsnewBF<-c()
FirearmsnewFP<-c()
FirearmsnewRank<-c()
FirearmsnewRank_BF<-c()

qt1<-subset(Firearms, select=c(BF,FP, Rank, Rank_BF))

FirearmsnewBF <- sapply(qt1$BF, function(x) if (x<150){paste("A",as.character((floor(x/5))*5),
                                sep="_")}else{if (x<200){
                                 x <- "A_150"}else{
                                  x <- "A_200"}})

FirearmsnewFP <- sapply(qt1$FP, function(x) if (x<150){paste("A",as.character((floor(x/5))*5),
                                sep="_")}else{if (x<200){
                                 x <- "A_150"}else{
                                  x <- "A_200"}})

FirearmsnewRank <- sapply(qt1$Rank, function(x)
 if (x<5) {"A_0"}else{
  if (x<10){paste("A",as.character((floor(x/5))*5), sep="_")}else{
   if (x<50){paste("A",as.character((floor(x/10))*10), sep="_")}else{
    if (x<100){paste("A",as.character((floor(x/25))*25), sep="_")}else{
     if (x<300){paste("A",as.character((floor(x/50))*50), sep="_")}else{
      if (x<1500){paste("A",as.character((floor(x/300))*300), sep="_")}else{
       "A_2450"}
     }
    }
   }
  }
 }
)
```

```
FirearmsnewRank_BF <- sapply(qt1$Rank_BF, function(x)
  if (x<5) {"A_0"}else{
    if (x<10){paste("A",as.character((floor(x/5))*5), sep="_")}else{
      if (x<50){paste("A",as.character((floor(x/10))*10), sep="_")}else{
        if (x<100){paste("A",as.character((floor(x/25))*25), sep="_")}else{
          if (x<300){paste("A",as.character((floor(x/50))*50), sep="_")}else{
            if (x<1500){paste("A",as.character((floor(x/300))*300), sep="_")}else{
              "A_2450"}
          }
        }
      }
    }
  }
)

Data<-subset(Firearms, select=c(Match, Make_DB, Model_DB, Make_Sample,
                                Model_Sample, Drag_Mark_Sample, Drag_Mark_DB,
                                Same_Model, Firing_Pin_Type_Sample,
                                Firing_Pin_Type_DB, Type_Sample, ActionLB_Sample, ActionLB_DB))

BF<-FirearmsnewBF
FP<-FirearmsnewFP
Rank<-FirearmsnewRank
Rank_BF<-FirearmsnewRank_BF

Data<-cbind(Data, BF, FP, Rank, Rank_BF)

# Define States
=============================================================================
st.Rank <- unique(FirearmsnewRank)
st.BF <- unique(FirearmsnewBF)
st.FP <- unique(FirearmsnewFP)
st.Match <- toupper(unique(Firearms$Match))
st.Make_DB <-toupper(unique(Firearms$Make_DB))
st.Model_DB <- toupper(unique(Firearms$Model_DB))
st.Make_Sample <- toupper(unique(Firearms$Make_Sample))
st.Model_Sample <- toupper(unique(Firearms$Model_Sample))
#st.IdentifierGun_Sample <- toupper(unique(Firearms$IdentifierGun_Sample))
#st.IdentifierGun_DB <- toupper(unique(Firearms$IdentifierGun_DB))
st.Primer_Sample <- toupper(unique(Firearms$Primer_Sample))
st.Primer_DB <- toupper(unique(Firearms$Primer_DB))
st.Drag_Mark_Sample <- toupper(unique(Firearms$Drag_Mark_Sample))
st.Drag_Mark_DB <- toupper(unique(Firearms$Drag_Mark_DB))
st.Rank_BF <- unique(FirearmsnewRank_BF)
st.Same_Model <- toupper(unique(Firearms$Same_Model))
```

```
st.Firing_Pin_Type_Sample <- toupper(unique(Firearms$Firing_Pin_Type_Sample))
st.Firing_Pin_Type_DB <- toupper(unique(Firearms$Firing_Pin_Type_DB))
st.Type_Sample <- toupper(unique(Firearms$Type_Sample))

st.ActionLB_Sample <- toupper(unique(Firearms$ActionLB_Sample))
st.ActionLB_DB <- toupper(unique(Firearms$ActionLB_DB))


# Create new Network
==================================================================
BN9MM <- CreateNetwork("BN9MM")
#DeleteNetwork(BN9MM)

NetworkTitle(BN9MM) <- "BN for the interpretation of 9MM results from IBIS"
NetworkComment(BN9MM) <- "KB Morris DoD grant"

# Create nodes
====================================================================
Rank <- NewDiscreteNode(BN9MM, "Rank", states=st.Rank)
BF <- NewDiscreteNode(BN9MM, "BF", states=st.BF)
FP <- NewDiscreteNode(BN9MM, "FP", states=st.FP)
Match <- NewDiscreteNode(BN9MM, "Match", states=st.Match)
Make_DB <- NewDiscreteNode(BN9MM, "Make_DB", states=st.Make_DB)
Model_DB <- NewDiscreteNode(BN9MM, "Model_DB", states=st.Model_DB)
Make_Sample <- NewDiscreteNode(BN9MM, "Make_Sample", states=st.Make_Sample)
Model_Sample <- NewDiscreteNode(BN9MM, "Model_Sample", states=st.Model_Sample)
#IdentifierGun_Sample <- NewDiscreteNode(BN9MM, "IdentifierGun_Sample",
states=st.IdentifierGun_Sample)
#IdentifierGun_DB <- NewDiscreteNode(BN9MM, "IdentifierGun_DB",
states=st.IdentifierGun_DB)
#Primer_Sample <- NewDiscreteNode(BN9MM, "Primer_Sample", states=st.Primer_Sample)
#Primer_DB <- NewDiscreteNode(BN9MM, "Primer_DB", states=st.Primer_DB)
Drag_Mark_Sample <- NewDiscreteNode(BN9MM, "Drag_Mark_Sample",
states=st.Drag_Mark_Sample)
Drag_Mark_DB <- NewDiscreteNode(BN9MM, "Drag_Mark_DB", states=st.Drag_Mark_DB)
Rank_BF <- NewDiscreteNode(BN9MM, "Rank_BF", states=st.Rank_BF)
Same_Model <- NewDiscreteNode(BN9MM, "Same_Model", states=st.Same_Model)

Firing_Pin_Type_Sample <- NewDiscreteNode(BN9MM, "Firing_Pin_Type_Sample",
states=st.Firing_Pin_Type_Sample)
Firing_Pin_Type_DB <- NewDiscreteNode(BN9MM, "Firing_Pin_Type_DB",
states=st.Firing_Pin_Type_DB)
Type_Sample <- NewDiscreteNode(BN9MM, "Type_Sample", states=st.Type_Sample)
```

```
ActionLB_Sample <- NewDiscreteNode(BN9MM, "ActionLB_Sample",
states=st.ActionLB_Sample)
ActionLB_DB <- NewDiscreteNode(BN9MM, "ActionLB_DB", states=st.ActionLB_DB)


# Change State Titles
=================================================================
NodeStateTitles(Rank) <- formatC(as.numeric(gsub("A_","",unique(Data$Rank))), width=4,
flag="0")
NodeStateTitles(BF) <- formatC(as.numeric(gsub("A_","",unique(Data$BF))), width=3,
flag="0")
NodeStateTitles(FP) <- formatC(as.numeric(gsub("A_","",unique(Data$FP))), width=3,
flag="0")
NodeStateTitles(Match) <- unique(Firearms$Match)
NodeStateTitles(Make_DB) <- unique(Firearms$Make_DB)
NodeStateTitles(Model_DB) <- unique(Firearms$Model_DB)
NodeStateTitles(Make_Sample) <- unique(Firearms$Make_Sample)
NodeStateTitles(Model_Sample) <- unique(Firearms$Model_Sample)
NodeStateTitles(Drag_Mark_Sample) <- unique(Firearms$Drag_Mark_Sample)
NodeStateTitles(Drag_Mark_DB) <- unique(Firearms$Drag_Mark_DB)
NodeStateTitles(Rank_BF) <- formatC(as.numeric(gsub("A_","",unique(Data$Rank_BF))),
width=4, flag="0")
NodeStateTitles(Same_Model) <- unique(Firearms$Same_Model)
NodeStateTitles(Firing_Pin_Type_Sample) <- unique(Firearms$Firing_Pin_Type_Sample)
NodeStateTitles(Firing_Pin_Type_DB) <- unique(Firearms$Firing_Pin_Type_DB)
NodeStateTitles(Type_Sample) <- unique(Firearms$Type_Sample)

NodeStateTitles(ActionLB_Sample) <- unique(Firearms$ActionLB_Sample)
NodeStateTitles(ActionLB_DB) <- unique(Firearms$ActionLB_DB)

# Add links
====================================================================
AddLink(Match, Rank)
AddLink(Match, Rank_BF)
AddLink(Match, FP)
AddLink(Match, BF)
AddLink(Match, Model_Sample)
AddLink(Match, Model_DB)

AddLink(Model_Sample, Model_DB)
AddLink(Model_Sample, Rank)
AddLink(Model_Sample, Rank_BF)
AddLink(Model_Sample, FP)
AddLink(Model_Sample, BF)
```

```
AddLink(Model_Sample, Make_Sample)
AddLink(Model_Sample, Firing_Pin_Type_Sample)
AddLink(Model_Sample, Drag_Mark_Sample)
AddLink(Model_Sample, Same_Model)
AddLink(Model_Sample, Type_Sample)
AddLink(Model_Sample, ActionLB_Sample)

AddLink(Model_DB, Rank)
AddLink(Model_DB, Rank_BF)
AddLink(Model_DB, FP)
AddLink(Model_DB, BF)
AddLink(Model_DB, Make_DB)
AddLink(Model_DB, Firing_Pin_Type_DB)
AddLink(Model_DB, Drag_Mark_DB)
AddLink(Model_DB, Same_Model)
AddLink(Model_DB, ActionLB_DB)

AddLink(Drag_Mark_Sample, Drag_Mark_DB)
AddLink(Firing_Pin_Type_Sample, Firing_Pin_Type_DB)
AddLink(ActionLB_Sample, ActionLB_DB)


# Add CPTs # (Conditional probability tables)
==================================================================
outfile <- tempfile("Data",fileext=".cas")
write.CaseFile(Data,outfile)
LearnCases(outfile,list(Rank, BF, FP, Match, Make_DB, Model_DB, Make_Sample,
                    Model_Sample, Drag_Mark_Sample, Drag_Mark_DB,
                    Rank_BF, Same_Model, Firing_Pin_Type_Sample,
                    Firing_Pin_Type_DB, Type_Sample, ActionLB_Sample, ActionLB_DB))

# Manipulate network
================================================================
CompileNetwork(BN9MM)

# Save the Network
==================================================================
SetNetworkAutoUpdate(BN9MM,TRUE)
WriteNetworks(BN9MM, "C:/Users/kbmorris.WVU-AD/Desktop/BN9MMJ15.dne")
#WriteNetworks(BN9MM, "C:/Users/Research Roger/Desktop/BN9MMJ15.dne")
#WriteNetworks(BN9MM, "C:/Users/KBM/Desktop/BN9MM.dne")
```

```
# LR Calc
===============================================================
Priors<-NodeBeliefs(Match)
PriorYes<-as.numeric(Priors[1])
PriorNo<-as.numeric(Priors[2])

# Lookup States  For No input assign e.g. BF.Score <- ""
BF.Score <- "A_125"
FP.Score <- "A_90"
TypeSampleState <- ""            #"PISTOL", "CARBINE", or "REVOLVER"
DragMarkSampleState <- "YES"          #"YES" or "NO"
DragMarkDBState <- ""              #"YES", "NO", or "UNKNOWN"
FiringPinTypeSampleState ="CIRCULAR"   #"CIRCULAR" or "GLOCK"
FiringPinTypeDBState =""       #"CIRCULAR" or "GLOCK"

lengthBF <- length(NodeStateTitles(BF))

ScoreDF <- c()
testa <- c()
for (i in 1:lengthBF){
  length(NodeStateTitles(FP))
  NodeFinding(BF) <- i
  NodeFinding(FP) <- i
  testa <- cbind(i, NodeFinding(BF),NodeFinding(FP))
  ScoreDF<-rbind(ScoreDF,testa)
}

ScoreDF <-as.data.frame(ScoreDF)
names(ScoreDF) <- c("Index", "BFlevel", "FPlevel")

RetractNodeFinding(BF)
RetractNodeFinding(FP)

if (BF.Score != ""){BF.pos <- grep(BF.Score,ScoreDF$BFlevel)
                    BF.pos.opt <- ScoreDF$BFlevel[BF.pos[1:length(BF.pos)]]==BF.Score
                    NodeFinding(BF) <- BF.pos[grep(TRUE, BF.pos.opt)]}

if (FP.Score != ""){FP.pos <- grep(FP.Score,ScoreDF$FPlevel)
                FP.pos.opt <- ScoreDF$FPlevel[FP.pos[1:length(FP.pos)]]==FP.Score
                NodeFinding(FP) <- FP.pos[grep(TRUE, FP.pos.opt)]}

if (TypeSampleState != ""){TypeSamples <- c("PISTOL", "CARBINE", "REVOLVER")
                   NodeFinding(Type_Sample) <- grep(TypeSampleState, TypeSamples)}

if (DragMarkSampleState != ""){DragMarks <- c("YES", "NO", "UNKNOWN")
```

```
                    DragMark.opt <- DragMarks==DragMarkSampleState
                    NodeFinding(Drag_Mark_Sample) <- DragMarks[grep(TRUE,
DragMark.opt)]}

if (DragMarkDBState != ""){DragMarks <- c("YES", "NO", "UNKNOWN")
                  DragMark.opt <- DragMarks==DragMarkDBState
                  NodeFinding(Drag_Mark_DB) <- DragMarks[grep(TRUE, DragMark.opt)]}

if (FiringPinTypeSampleState != ""){FiringPinTypes <- c("CIRCULAR", "GLOCK")
                        NodeFinding(Firing_Pin_Type_Sample) <-
grep(FiringPinTypeSampleState, FiringPinTypes)}

if (FiringPinTypeDBState != ""){FiringPinTypes <- c("CIRCULAR", "GLOCK")
                     NodeFinding(Firing_Pin_Type_DB) <- grep(FiringPinTypeDBState,
FiringPinTypes)}


if ((NodeFinding(BF)==BF.Score)&(NodeFinding(FP)==FP.Score)){cat("Good to
go")}else{cat("Problem", NodeFinding(BF), NodeFinding(FP), "\n")}

Posteriors<-NodeBeliefs(Match)
PosteriorYes<-as.numeric(Posteriors[1])
PosteriorNo<-as.numeric(Posteriors[2])

LR <- (PosteriorYes*PriorNo)/(PosteriorNo*PriorYes)
LLR <- log10(LR)
LR
LLR

NodeFinding(BF)
NodeFinding(FP)
NodeFinding(Type_Sample)
NodeFinding(Drag_Mark_Sample)
NodeFinding(Drag_Mark_DB)
NodeFinding(Firing_Pin_Type_Sample)
NodeFinding(Firing_Pin_Type_DB)

# Reset Network
=================================================================
RetractNodeFinding(BF)
RetractNodeFinding(FP)
RetractNodeFinding(Type_Sample)
RetractNodeFinding(Drag_Mark_Sample)
RetractNodeFinding(Drag_Mark_DB)
RetractNodeFinding(Firing_Pin_Type_Sample)
```

```
RetractNodeFinding(Firing_Pin_Type_DB)

RetractNetFindings(BN9MM)

#Finish
==================================================================
DeleteNetwork(BN9MM)


==================================================================
==================================================================
# All LR Calc
==================================================================
Priors<-NodeBeliefs(Match)
PriorYes<-as.numeric(Priors[1])
PriorNo<-as.numeric(Priors[2])
all.LR <- c()
tmp.LR <-c()

for (kk in st.BF){
  for (ll in st.FP){

# Lookup States
BF.Score <- kk
FP.Score <- ll

lengthBF <- length(NodeStateTitles(BF))

ScoreDF <- c()
testa <- c()
for (i in 1:lengthBF){
  length(NodeStateTitles(FP))
  NodeFinding(BF) <- i
  NodeFinding(FP) <- i
  testa <- cbind(i, NodeFinding(BF),NodeFinding(FP))
  ScoreDF<-rbind(ScoreDF,testa)
}

ScoreDF <-as.data.frame(ScoreDF)
names(ScoreDF) <- c("Index", "BFlevel", "FPlevel")

BF.pos <- grep(BF.Score,ScoreDF$BFlevel)
BF.pos.opt <- ScoreDF$BFlevel[BF.pos[1:length(BF.pos)]]==BF.Score

FP.pos <- grep(FP.Score,ScoreDF$FPlevel)
```

```
FP.pos.opt <- ScoreDF$FPlevel[FP.pos[1:length(FP.pos)]]==FP.Score

NodeFinding(BF) <- BF.pos[grep(TRUE, BF.pos.opt)]
NodeFinding(FP) <- FP.pos[grep(TRUE, FP.pos.opt)]

if ((NodeFinding(BF)==BF.Score)&(NodeFinding(FP)==FP.Score)){cat("Good to
go")}else{cat("Problem", NodeFinding(BF), NodeFinding(FP), "\n")}

Posteriors<-NodeBeliefs(Match)
PosteriorYes<-as.numeric(Posteriors[1])
PosteriorNo<-as.numeric(Posteriors[2])

LR <- (PosteriorYes*PriorNo)/(PosteriorNo*PriorYes)
LLR <- log10(LR)
cat(kk,ll,LR,LLR,"\n")

tmp.LR <- cbind(kk,ll,LR,LLR)
all.LR <- rbind(all.LR,tmp.LR)

RetractNetFindings(BN9MM)
  }
}
all.LR <- as.data.frame(all.LR)
names(all.LR) <- c("BFlevel", "FPlevel", "LR", "LLR")
write.csv(all.LR,"Z:/BN9MM_AllLRs.csv", row.names=FALSE)
```