Document Title:	A Study of False-Positive and False-Negative Error Rates in Cartridge Case Comparisons
Author(s):	David P. Baldwin, Stanley J. Bajic, Max Morris, and Daniel Zamzow
Document No.:	NCJ 249874
Date Received at NCJRS:	May 2016
Contract Number:	DE-AC02-07CH11358

A Study of False-Positive and False-Negative Error Rates in Cartridge Case Comparisons

David P. Baldwin, Stanley J. Bajic, Max Morris, and Daniel Zamzow

Ames Laboratory, USDOE

Technical Report # IS-5207

April 7, 2014

DFBA

DEFENSE FORENSICS & BIOMETRICS AGENCY

Office of the CHIEF SCIENTIST

DEFENSE FORENSIC SCIENCE CENTER

Sponsors and Disclaimer

This work was supported by Defense Biometrics and Forensics Office, Assistant Secretary of Defense (Research and Engineering), through the U.S. Department of Energy Contract No. DE-AC02-07CH11358. 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 Defense Biometrics and Forensics Office, 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.

Acknowledgements

This work was supported by Defense Biometrics and Forensics Office through the U.S. Department of Energy under Contract No. DE-AC02-07CH11358. The study would have been impossible without the willing participation of the hundreds of examiners and the permission of their agencies to participate voluntarily. The study was managed by the Defense Forensics and Biometrics Agency and the Defense Forensic Science Center, Office of the Chief Scientist. Henry Maynard, Garold Warner and other members of the OCS staff located at the Defense Forensic Science Center shaped and guided the project. We give special thanks to Dr. Michael "Jeff" Salyards, Executive Director, DFSC, Rick Tontarski, DFBA /DFSC Chief Scientist, and the staff of the firearms section at USACIL, a division of the DFSC. The Forensic Initiative at WVU provided the firearms, ammunition, and support in firing and collecting samples for the study including Dr. Keith Morris, Mike Bell, and several student staff members. The Forensic Research Committee of ASCLD provided input and guidance as well as valuable recruitment support for both the pilot study and the full study and we want to particularly thank Jay Henry for his efforts in communications through ASCLD membership. AFTE provided the largest group of participants through the support and communications provided by Jay Stuart. The Story County, Iowa Sheriff's Office made the pilot study possible and we want to particularly thank Sheriff Fitzgerald, Sergeant Backous, and Detective Rhoads. SWGGUN Chair Andy Smith provided valuable guidance in designing the study and aided in identifying the proper selection of ammunition. The staff and management of the Firearms & Toolmarks Unit at the FBI Laboratory provided additional input on project design. Finally we thank the staff of the Ames Laboratory Shipping and Receiving Department for coordinating and handling shipment of all of the materials for the study, in particular Vicki Sieve.

Abstract: This report provides the details for a study designed to measure examiner (not laboratory) error rates for false identifications and false eliminations when comparing an unknown to a collection of three known cartridge cases. Volunteer active examiners with Association of Firearm and Toolmark Examiners (AFTE) membership or working in laboratories that participate in ASCLD were provided with 15 sets of 3 known + 1 unknown cartridge cases fired from a collection of 25 new Ruger SR9 handguns. The ammunition was all Remington 9-mm Luger (manufacturer designation L9MM3) and sets were made up of cartridge cases fired within 100 cartridges of each other for each gun. During the design phase of the experiment, examiners had expressed a concern that known samples should not be separated by a large number of fired cartridges. However, studies published on this effect indicate that several thousands of cartridges could be fired by the same firearm without making the identifying characteristics change enough to prevent identification. [1] Examiners were provided with a background survey, an answer sheet allowing for the AFTE range of conclusions, and return shipping materials. They were also asked to assess how many of the 3 knowns were suitable for comparison, providing a measured rate of how often each firearm used in the study produces useable, quality marks. The participating examiners were provided with known positives and known negatives from independent groups of samples, providing independent measurements of a false-positive rate and independent measurements of a false-negative rate, allowing the study to measure both rates and uncertainties in those rates.

Responses were received from 218 participating examiners. The rate of false negatives (estimated as 0.367% from comparisons known to be from the same firearm but reported as eliminations) was quite low with the error distributed across examiners of various backgrounds (state, federal, local, private, etc. as determined from self-reported survey information). The overall rate of false positives (estimated as 1.01% from comparisons known to be from different firearms but reported as identifications) was significantly higher. However, most of the errors were reported by a small number of examiners; that is, individual examiners have varying error rates. For most examiners this is quite low while for some it is relatively high. Hence the overall rate is best interpreted as an average of widely varying individual rates. Inconclusive results were not recorded as errors. Rates of poor quality mark production for these handguns varied across the 25 sample handguns. Those rates were 2.3 (±1.4) %.

False-positive and false-negative error rates for individual examiner performance on comparisons were measured. The rates are not uniform across the sample population with a few examiners providing most of the false-positive responses. False-negative rates are low and comparable to or lower than the rate of production of poor quality marks by the firearms used in this study. Laboratory error rates may be significantly lower than these individual rates if quality assurance procedures are applied that can effectively manage to reduce or eliminate the propagation of false positives reported by individuals.

Introduction:

This study was designed to provide a better understanding of the error rates associated with the forensic comparison of fired cartridge cases. Several previous studies have been carried out to examine this and related issues of individualization and durability of marks [1-5], but the design of these previous studies, whether intended to measure error rates or not, did not include truly independent sample sets that would allow the unbiased determination of false-positive or false-negative error rates from the data in those studies.

The Admissibility Resource Kit (ARK) developed and published by the Scientific Working Group for Firearms and Toolmarks (SWGGUN) maintains an extensive bibliography of literature relevant to the reliability, repeatability, and validation of forensic examinations of cartridge cases. Currently, this bibliography can be found at:

http://www.swggun.org/swg/index.php?option=com_content&view=article&id=5:testability-of-the-scientific-principle&catid=9:ark#CC

Please note that there are tentative plans to transfer the maintenance of the ARK to AFTE in the event that SWGGUN is replaced with another standards organization.

The study was designed with sample sets for comparison that are as independent as economically feasible given the cost of firearms and ammunition. We set out to measure both false-positive rates and false-negative rates, so the participants were presented with 15 independent comparison sets that we knew either came from the same source or didn't. No source firearm was repeated within any participant's test packet except within a set that was from the same source. This was deemed important since there was anecdotal feedback during discussions of design that examiners remember patterns even outside a posed comparison, and it was reported that this might affect the responses or the perception of the experimental design. Initially we had considered designing each set as a 1 to 1 comparison. We obtained feedback that some labs do not allow this type of analysis so we decided instead to provide examiners with 3 "known" cartridge cases to compare to 1 "questioned" case. This was meant to mimic the situation where an examiner would compare a questioned case to repeated firings from a firearm in evidence. The provision of 3 knowns in this study allowed us to address an issue of determining how often each firearm in the study would produce a fired cartridge case with insufficient markings for a comparison. In each set, the participants were asked first to determine how many of the 3 knowns had marks that were suitable for comparison. One valuable outcome of this study is an extensive measurement of this poor mark production rate for the 25 SR9 handguns used in this study. It is important to remember that these poor marking rates might well be different with different models of firearm, firearms of significantly different age or condition, and different makes and manufacture of ammunition. However, our data provides a reliable measurement of this phenomenon for the samples used in this study. The measurement of this poor marking rate allows us to understand the potential effects on the measured examiner error rates. By measuring this rate of poor mark reproduction, we also avoid a problematic practice of prescreening the quality of samples provided to the participants. In some previous studies, a qualified examiner would screen all samples to make sure that only well-marked samples were included. This introduces two problems. One is that the study risks the criticism that only "easy" samples are included, which is not necessarily reflective of real casework samples. The second is that even if the screener is very qualified (and can afford or has the fortitude to examine the 20,000 cases used in this study) the judgment of quality marks is a subjective judgment best left to each examiner. By instead allowing all of the examiners to report a rate of poor production for their known samples we gain an insight into the effect without having to accept the cost and ambiguity of a reanalysis for every sample.

A very important aspect of this work that needs to be clearly understood is that the study specifically asked participants not to use their laboratory or agency peer review process. There are two aspects to consider in this decision. First, the errors that are identified during the study and the estimated rates of error are for individual examiner performance and not directly related to rates of error for reported analyses from any laboratory or agency. The variability in how, how often, and how effectively the quality assurance programs in the dozens of agencies in which the participants practice might have detected and prevented reporting errors is so wide and undefined that we deemed a study that included these QA measures as being less meaningful than a study of how well trained individuals perform in a relatively uniform set of circumstances. After reviewing the results of this study, we would hope that QA managers would institute appropriate controls to manage the measured performance error rates, rather than taking a snapshot of the systemic performance of existing disparate systems without understanding what errors are being managed by those systems.

The firearm chosen for this study is the Ruger SR9 semiautomatic 9-mm handgun. This is a relatively new model of handgun, but it was chosen for several reasons. Sturm, Ruger & Company is a popular firearms brand with a reasonably positive reputation for service and quality. Although not the most popular source for semiautomatic handguns that might be chosen as a service sidearm compared to industry giants like Glock, Sig Sauer, Beretta, and S&W, Ruger has an apparent sizeable share of the market for home defense weapons. In part this is due to the lower cost of their handguns compared to similarly featured products from these other manufacturers. At the time we began this study, the retail prices for the Ruger SR9 pistol were as much as 25 to 30% lower than a comparable 9-mm Glock pistol, making a study using the Ruger SR9 more affordable. For this same reason, Ruger products may be more likely to be obtained and available for illegal activities. Currently, the Ruger P95 is a more mature product line that exists in larger numbers than the Ruger SR9. However, Ruger was discontinuing production of their P-series firearms, and have since halted production of the P95 (in October 2013). The Ruger SR series is fully replacing the P series as their full size semiautomatic handgun line. We would expect the SR series to eventually rival if not surpass numbers of P-series firearms in circulation. An additional reason for choosing the Ruger SR9 was the need for reliable performance for production of the quantity of samples needed for this study. The striker-fired design, similar to that employed by the popular Glock line, has proven to be quite reliable compared to other hammer-based designs and uses fewer moving parts. While there were several anecdotal suggestions that cartridge cases from firearms like the Glock might have been "too easy" or the Bryco/Jennings/Jimenez style handguns might be "too hard" or have too high a rate of poorly reproduced marks, the Ruger SR9 had no such preconceived biases for or against its use. This study incorporates a design to measure the reproduction rate for useable comparison marks, and the relative ease of comparison is an issue that will always need to be addressed to generalize this type of study to more than one model of firearm. We chose to base the available responses on the AFTE Range of Conclusions, given the preponderance of AFTE members as participants in the study. Therefore, we interpret the meaning of inconclusive according to the AFTE guidelines that establish this range of conclusions (http://afte.org/AssociationInfo/comm & info/roc.htm – also reproduced below). If the examiner does not find sufficient matching detail to uniquely identify a common source for the known and questioned samples, and there are no class characteristics such as caliber that would preclude the cases as having been fired from the same-source firearm, a finding of inconclusive is an appropriate answer (and not counted as an error or as a non-answer in this study). The underlying rationale for this finding of inconclusive is that the examiner is unable to locate sufficient corresponding individual characteristics to either include or exclude an exhibit as having been fired in a particular firearm and the possible reasons are numerous as to why insufficient marks exist. As is determined in this study, there are also a significant number of times that the firearm fails to make clear and reproducible marks (which very well might have happened for a questioned case). On the other hand, we received many conflicting comments from study participants regarding the use of this range of conclusions, indicating systemic differences in how closely these guidelines are followed or how the guidelines are interpreted. Some participants indicated that they would not participate if the full range of conclusions, including different classifications of conclusions for inconclusive findings, were not included. Other participants indicated that their agencies did not allow them to distinguish between the different classifications for inconclusive results so they chose not to differentiate between any of the three options within the AFTE guidelines in their answers. Some indicated that the design of our study with all cartridges fired from the same model of firearm using the same type of ammunition would prohibit the use of a finding of elimination, while others used a mixture of inconclusive and elimination or did not use inconclusive at all to indicate a finding other than identification. In this report we present our findings on these variations and suggest that this is an area for further study and discussion within the community.

One of the reasons for undertaking this work was to address an issue identified in the 2009 National academies report: "Strengthening Forensic Science in the United States: A Path Forward."[6] In the summary assessment section on toolmarks and firearms identification in that report the authors stated that: "Sufficient studies have not been done to understand the reliability and repeatability of the methods." This work addresses that issue through an assessment of the reliability of the comparison of fired cartridge casings.

AFTE Range of Conclusions, reproduced from http://afte.org/AssociationInfo/comm & info/roc.htm.

The examiner is encouraged to report the objective observations that support the findings of toolmark examinations. The examiner should be conservative when reporting the significance of these observations.
The following represents a spectrum of statements:

1) IDENTIFICATION

Agreement of a combination of individual characteristics and all discernible class characteristics where the extent of agreement exceeds that which can occur in the comparison of toolmarks made by different tools and is consistent with the agreement demonstrated by toolmarks known to have been produced by the same tool.

2) INCONCLUSIVE

A. Some agreement of individual characteristics and all discernible class characteristics, but insufficient for an identification.

B. Agreement of all discernible class characteristics without agreement or disagreement of individual characteristics due to an absence, insufficiency, or lack of reproducibility.

C. Agreement of all discernible class characteristics and disagreement of individual characteristics, but insufficient for an elimination.

3) ELIMINATION

Significant disagreement of discernible class characteristics and/or individual characteristics.

4) UNSUITABLE

Unsuitable for examination.

Finally, the study was limited to participants who are self-reporting as active forensic firearm examiners who are either AFTE members or work in laboratories that participate in ASCLD. Although it might be desirable to understand how non-practicing or untrained participants might perform under the same circumstances as trained examiners, there are important statistical reasons for not including trainees. The expected rates of error are low enough that dividing our participant pool into subgroups that are trained and not trained would add cost to the study without adding enough participants to allow a precise measurement of error rates for this group of trainees. It was deemed more important to measure the error rates for trained practicing examiners accurately and precisely than to measure the effect of another variable with much less precision and accuracy.

Experimental:

Pilot study:

Before embarking on soliciting participants and assembling test materials for the full study, we conducted a pilot test to solicit input from volunteers from four laboratories recruited through the ASCLD Forensic Research Committee (FRC). Ammunition and firearms purchases for this study were delayed for several months by contractual arrangements and ongoing market supply issues at the time. As a result, test samples were obtained through the voluntary cooperation of a local police agency. The test sets (240 fired cases) were collected by the Story County, Iowa Sheriff's Department using five of their service sidearms (.40 S&W caliber, Sig Sauer P229), obtained during a training session for that department. Four pilot test cartridge case packets were assembled and sent to volunteers at the four volunteer agencies. We requested feedback on the study design, survey forms, informed consent forms, and any other information they wanted to provide. Several improvements to the study materials were made as a result of this pilot study. One significant limitation of the pilot study was the limited number of firearms available to assemble the test sets, resulting in many more repeated source samples than were used in the full study.

Test firing and Selection of Ammunition:

Among those consulted during the design phase of this study was Andrew Smith, SWGGUN Chair and a firearms examiner with the San Francisco Police Department. He encouraged us to test fire several types of ammunition with the model of handgun to be used in the study. We had obtained three test fired samples of six commonly available 9-mm cartridges from Blazer, Winchester (two types), Federal, Remington, and PMC and sent them to Mr. Smith for evaluation. All cartridges produced useable detail.

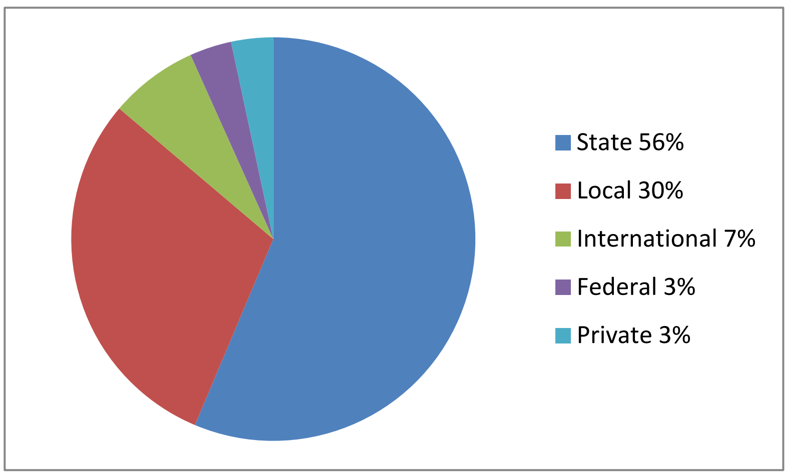

Figure 1: Distribution of the 284 respondents enrolled in the study.

Solicitation of Participants:

Due to the use of human subjects in this study, the experimental program was subject to review and approval by Institutional Review Boards at Ames Laboratory (using the Institutional Review Board of our contracting agency, Iowa State University) and by the Department of Defense. The study was designed to protect the participants from risk to their professional standing and reputation by making all results strictly anonymous. One group of researchers prepared materials and scored the responses and a separate group solicited participants and sent and received sealed unlabeled study material packets. No researcher has information connecting identities with scored results.

Invitations to participate in this study were broadcast by email and newsletter publications through ASCLD and AFTE. The invitation letter and the informed consent form for this study are included in Appendix A. The study accepted registrations to participate during the months of August and September of 2013. We received consent forms and issued test materials to 284 participants and received completed results from 218 respondents by the end of December, 2013. Answers and identities, including significant identifying information like home organization, have been completely separated. However, from our respondent pool we can determine that 56% of those enrolled were employed by domestic state-level agencies, 30% were employed by domestic local agencies, 7% were employed by international (non-U.S.) agencies, 3% were employed by domestic federal agencies, and 3% were self-employed or worked for private forensic agencies.

Figure 2: A Ruger© SR9© handgun with two 17-round magazines. This is the same model as the firearms used in this study.

Sample Preparation and Collection:

Given restrictions on acquisition and property management of firearms by the Ames Laboratory, our collaborators at the Forensic Science Initiative at West Virginia University (WVU) purchased twenty-five (25) new Ruger SR9, 9mm Luger centerfire pistols. The serial numbers and assigned identifiers (explained below) used for each handgun in this study are listed in the Table in Appendix B. WVU also purchased 25,000 cartridges of Remington UMC Pistol and Revolver Cartridges. The cartridges used were 115 grain full metal jacket (FMJ) bullets. The ammunition came in boxes of 50 cartridges with each box containing a 50-holed plastic bullet carrier to keep the individual cartridges from “bouncing around” during shipping and handling, potentially acquiring additional marks. The ammunition came from two lots from the manufacturer: 136 boxes from Lot# 3503 ODEL 2195989 and 264 boxes from Lot# 350 ODFL 32206416. Prior to the sample collection campaign for this study, staff from WVU fired 200 cartridges of ammunition using each handgun and collected the fired cases for use in an unrelated study. There is anecdotal evidence to suggest that there is some rapid early change in marks made in the first several firings with a new handgun. The cartridges were collected by WVU to investigate this effect in a separate study. For the same reason, these fired cartridge cases were not used in this error-rate study.

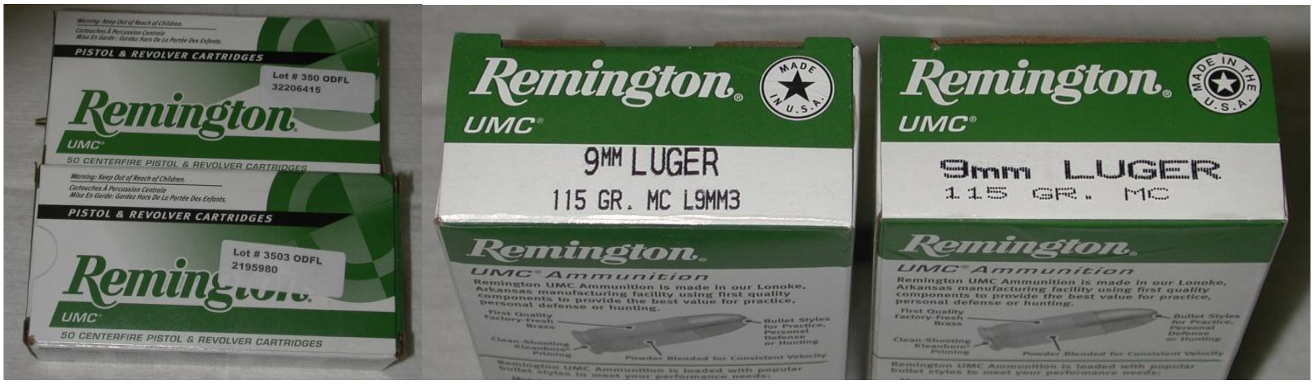

Figure 3. Boxes containing ammunition used in this study.

The ammunition was divided up into batches of 16 boxes, which correspond to 800 cartridges per batch. Each batch was allotted to one of the 25 Ruger SR9 handguns. Except for handgun “B1,” each handgun was used to fire all 800 cartridges of ammunition from the same Lot. Handgun “B1” was used to fire 400 cartridges of ammunition from each of the two lots.

The sample sets for the study were made up of 15 comparisons of 3 knowns to 1 questioned cartridge case. For all participants, 5 of the sets were from known same-source firearms, and 10 of the sets were from known different-source firearms. We instructed all participants to refrain from sharing or discussing the contents or results of their sample sets and answers to minimize the risk of revealing this design. We did not share these details to anyone outside the group assembling the test sets and the local project managers. In order to provide 5 independent sets of samples to measure a rate of false-negative conclusions, which could then be used to determine not only a rate but an uncertainty in that rate, the 25 handguns were broken into 5 groups of 5 handguns (A1-A5, B1-B5, C1-C5, D1-D5, E1-E5). Each participant was randomly assigned to one of the groups A through E. If assigned to group A, that examiner would receive 5 sets where cartridges from handgun A1 would be used for all of the cartridge cases in one set through A5 making up the fifth set of all same-source cartridge cases. These designations did not appear on any markings provided to the participants and the order within the 15 sets within a test packet was randomized. The actual identity of the sources and the assembly of the test kits and was tracked by an individual on the project team without knowledge of the identity of the participants, and another individual checking the accuracy of his work (also not privy to the identity of the participants). The remaining 10 sets for each examiner were assembled from different-source firearms from different groups. Those in the A group were also provided 5 sets with firearms from the B group producing the knowns and D firearms producing the questioned cartridge cases, and 5 sets with firearms from C producing the knowns and E firearms producing the questioned cartridge cases. In order to further randomize the sets between examiners in a given group, the pair of firearms in the different-source comparison sets was incremented for each examiner so that only every sixth participant in a group would receive the same comparison group. So the A group of examiners saw A(k) v. A(q), B(k) v. D(q), and C(k) v. E(q) (likewise the remaining groups were assigned as B group: B-B/C-E, D-A; C group: C-C/ D-A, E-B; D group: D-D/ E-B, A-C; E group: E-E/ A-C, B-D). The net effect of this design is that every participant received cartridges from each of the 25 firearms as either a known or questioned sample (or both for known same-source sets), each known same-source group was independent of the other four groups, and no firearm was used more than once for a participant unless it was part of a known same-source set.

Table I: Experimental Set Design

Group	A	B	C	D	E
Same-source comparisons	A1-A1	B1-B1	C1-C1	D1-D1	E1-E1
	A2-A2	B2-B2	C2-C2	D2-D2	E2-E2
	A3-A3	B3-B3	C3-C3	D3-D3	E3-E3
	A4-A4	B4-B4	C4-C4	D4-D4	E4-E4
	A5-A5	B5-B5	C5-C5	D5-D5	E5-E5
Different-source comparisons	B v D: 1v2, 2v3, 3v4, 4v5, 5v1 and other skip permutations	C v E	D v A	E v B	A v C
	C v E	D v A	E v B	A v C	B v D

False-negative error rates for each of the 5 same-source groups (A through E in Table I) provide 5 independent measurements of this rate and allow the measurement of uncertainty in the false-negative rate. False-positive error rates for each of the 5 different-source groups (B v. D, C v. E, D v. A, E v. B, and A v. C in Table I) provide 5 different independent measurements of the false-positive rate and allow the measurement of uncertainty in the false-positive rate.

Cartridge cases were collected in groupings of 100 with 800 total cartridge cases collected per firearm. From a batch for a particular firearm, 100 cartridges of ammunition were loaded into six magazines; five magazines loaded with 17 cartridges and one magazine loaded with 15 cartridges. Loading of the magazines took place in the observation area of the firing range. Once loaded, the six magazines were taken into the firing range and used. The handgun was discharge through a homemade brass catcher (Figure 4) with the cartridge cases collected by the catcher after ejection from the firearm. After firing all 100 cartridges, the cartridge cases were removed from the catcher, placed in a sealable plastic bag, and then taken out of the firing range and into another section of the observation area. The cartridge cases were then placed in 100 round plastic ammunition carrier boxes to prevent any additional marks being made on the cases and to facilitate sorting and transporting.


Figure 4. Brass catcher used to collect samples for this study.

The ammunition was fired and collected sequentially (i.e., 1-100, 101-200… 701-800). Only cartridge cases caught by the catcher were collected. If a cartridge case somehow fell out of the catcher, it was immediately discarded. Misfires were cleared from the firearm and disposed of without attempting to reuse the round. The 100-round ammunition carrier boxes were labeled with the firearm letter designation and the segment of cartridges that were collected (e.g., “Gun C2, 501-600”).

The firearms were cleaned prior to being used at the range and cleaned at the range after 400 cartridges were fired through the handgun. Cleaning consisted of disassembling the handgun, letting it cool; brushing the breach face with a plastic bristled brush; and running a cotton bore cleaning swab, wetted with Hoppe's No. 9, through the barrel. The firearm was then reassembled and either used to fire the remaining ammunition for that handgun or placed in its carrying case for storage.

Each collected cartridge case was labeled with a unique alphanumeric identifier. The alphanumeric identifier had the format of either “Kmfrcxxxyyy” or “Qmfrcxxxyyy”, where “x” is alphabetic letter and “y” is an integer. “K” and “Q” indicate that the labeled cartridge case would be used as a “known” or “questioned” casing in the generated sample sets. This was included so that if a participant were to mix up the K’s and Q’s within a set, they could still be identified. The 20,000 unique alphanumeric identifiers were randomly generated using a free online random-code generator (http://www.mindflow.com.au/Custom_CMS/index.php/Online-Tools/Random-Code-Generator-Free-Tool.html). The output from the random-code-generator was a csv format file that could be imported into a label making program.

Labels for the casings were generated using a Brother P-touch PT-2430PC Label Maker on ¼” white tape with black print (TZe-211) that was laminated. The printed labels were then manually affixed to each individual cartridge case. Figure 5 depicts examples of labeled cartridge cases. Although the labels as affixed have in some instances covered the ejector marks on the cartridge cases, the participants could have removed this label if necessary to make a definitive examination. There is some possibility that participants may have felt restrained from doing so for some reason and may have performed the examinations without information they normally would have used. We received only one comment to this effect. The labels were placed in this way due to the difficulty, time, and expense involved in placing the labels in a less obtrusive location, such as inside the cartridge cases. Had the study been significantly smaller (with fewer samples to label, to assemble into sets, and check for accuracy in assembly), we would have considered different placement options.

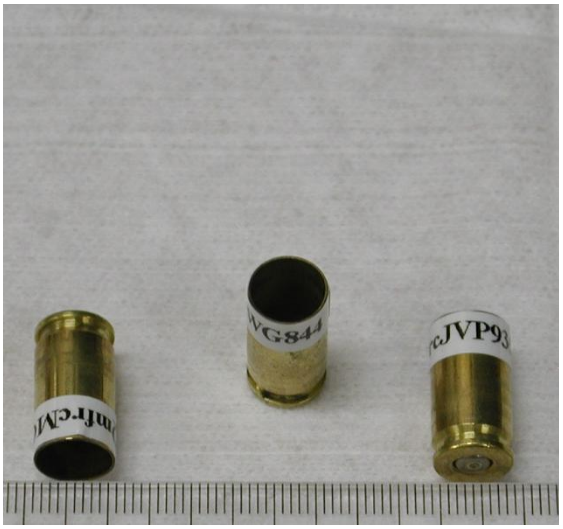

Figure 5. Labelled cartridge cases.

The sequence of the sets in each group was determined by a random number/list generator so that the order of same-source and different-source sets was random for each participant. The generator can be found at (http://www.random.org/lists/).

The response form is shown in Appendix C. Each form was marked with an identifier to associate it with the materials provided so that responses could be compared to known sources. The form includes background demographic information, the full range of AFTE-approved responses, and an evaluation of how many of the knowns in each set were suitable for comparison.

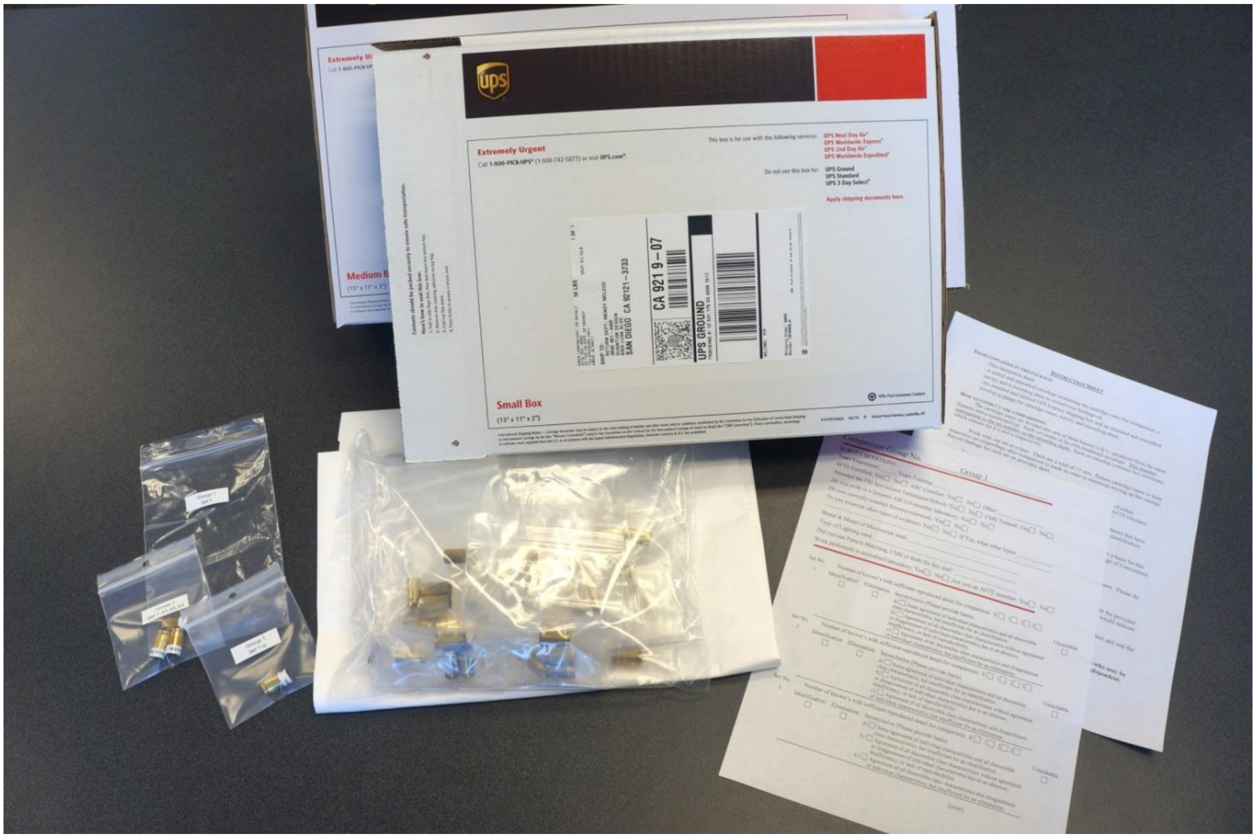

Figure 6. A sample package with supplied samples, return packaging, and forms.

Sample packages were assembled for each participant and included an instruction sheet (see Appendix D), a survey and answer sheet (see Appendix C), a return prepaid shipping box, an unmarked Tyvek envelope for returning the answers and samples, and plastic zip seal bags containing the sample sets marked with a set number (with knowns and questioned cartridge cases also in separate zip bags). The answer sheet and samples were marked with an identifier so that they could be recorded and scored upon return. The shipping box had a separate unrelated identifier that was unassociated with the sample set but connected to the respondent's identity. When materials were returned, that identifier was used to note that a participant was finished and should be removed from reminder mailing lists. The returned samples and answer sheets, sealed in the unmarked Tyvek envelope, were transferred to the scoring group – at which point the identity of the respondent would be separated from the results of the examination.

Several of the participants in the study were working outside the United States. In order to send sample sets to these participants and comply with U.S. export control and arms control regulations, the samples needed to be altered in a way that would prevent reloading of the fired cartridge cases, while still permitting examination. For these sample sets, each sample was cut with an approximately 3-mm long cut along the case wall using a handheld rotary tool with a cutting wheel. Examples of these altered cartridge cases are shown in Figure 7.

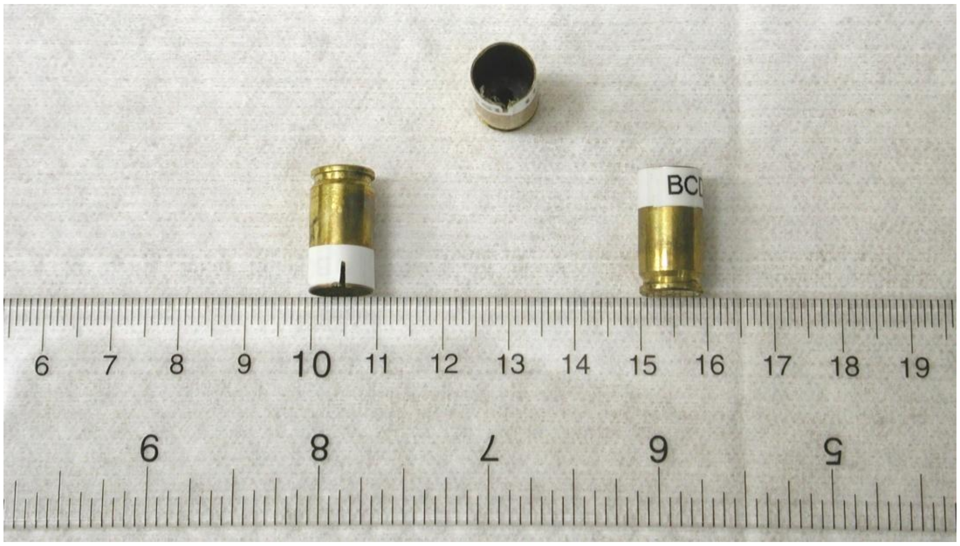

Figure 7. Sample cartridge cases intentionally altered for shipment to international participants.

Results and Discussion:

Of the 1090 true same-source comparisons made (i.e. where the three knowns and the single questioned samples actually were produced by the same firearm), only four comparisons were labeled elimination and eleven were judged inconclusive. The proportion of false elimination calls, relative to all responsive answers (elimination, identification, and inconclusive) was 4/1090 = 0.003670, or 0.3670%. The Clopper-Pearson exact 95% confidence interval [7] for this proportion is (0.001001, 0.009369), or (0.1001%, 0.9369%). Two of the four incorrect elimination calls were made by the same examiner; 215 of the 218 examiners did not make any false elimination calls. The proportion of false elimination and inconclusive calls together (not all errors, but also not identifications), relative to all responsive calls, was 15/1090 = 0.01376, or 1.376%; the corresponding 95% confidence interval is (0.007722, 0.02260), or (0.7722%, 2.260%). For those unfamiliar with statistical methods for confidence interval calculation, please note that Clopper-Pearson is a widely used method of calculating the confidence interval for a true value of a mean from sample data. Using a different method for confidence interval calculation would not affect the reported mean value of the error rate and would result in only minor changes in the size of the range of the confidence interval. Comparing these differences in confidence interval calculations is irrelevant to the discussion or the implications of this report.

One relevant number for comparison to the false elimination probability is the proportion of known samples that were judged to be poorly marked, and so not used by the examiners. Of the 3234 comparisons for which the number of useable knowns was reported (some examiners either did not provide this information or provided it for only some of the samples), all three specimens were used in 3018 cases, two were used in 207 cases, and only one was used in nine cases. Hence the "raw" proportion of known specimens judged by the examiners to be inappropriate for inclusion in the comparison was 225/9702 = 0.02319, or 2.319%, with a corresponding 95% confidence interval of (0.02174, 0.02827), or (2.174% to 2.827%). This is substantially greater than the estimated proportion of false eliminations, and raises the possibility that at least some of these incorrect eliminations may have been the result of poorly marked questioned samples. In addition, compared to the rate of false eliminations and inconclusive results for same-source samples, above, this rate of poorly marked cases may very well explain most if not all of these calls. At best this rate of poor marking makes determination of a false-negative error rate difficult since it is at least comparable in magnitude and indistinguishable in result.

Of the 2180 true different-source comparisons possible (i.e. where the three knowns and the single questioned sample actually were not produced by the same firearm), 22 comparisons were labeled identifications. The remaining correct responses included 735 reported as inconclusive and 1421 as eliminations. The proportion of false identification calls from among all responsive answers was 22/2178 = 0.01010, or 1.010% (two comparisons were not reported or left blank – reducing the total responses from 2180 to 2178). All but two of the 22 false identification calls were made by five of the 218 examiners, strongly suggesting that this error probability is not consistent across examiners (or in effect, that each examiner has his or her own false identification probability, and that these probabilities vary substantially).

As a result, appropriate statistical modeling for the proportion of false identifications is more intricate because it cannot be assumed that the probability of such a call is uniform across the participating examiners (and so the dataset). The beta-binomial model was used as a basis for estimation here; it is a mixture of binomial distributions (appropriate for each individual examiner) in which the individual binomial probabilities follow a beta distribution (representing the variation among examiners). Because the individual error probabilities can only be estimated with substantial uncertainty from the small sample of comparisons made by each examiner, the inference is less precise than the Clopper-Pearson methodology used for false eliminations. The maximum likelihood estimator for the false identification probability, averaged across examiners, is 0.00939 (0.939%), with a likelihood-based 95% confidence interval of (0.00360, 0.02261) or (0.360%, 2.261%).

Examiners were also highly heterogeneous in the number of true different-source cartridge cases they identified as inconclusive. Of 218 examiners, 96 (the largest subcategory) labeled none of the comparisons as inconclusive (instead using elimination to denote a determination that found insufficient matching detail for an identification), while 45 (the second largest subcategory) labeled all 10 of the comparisons as inconclusive. The remaining 77 examiners were fairly evenly spread between these two extremes, with somewhat more of them reporting relatively fewer inconclusive calls. There are mild inverse correlations between the number of inconclusive/nonresponse calls made with the known different-source cases, and the reported number of years of training (correlation = -0.1393) and number of years of experience (correlation = -0.1034); that is, there is a weak tendency for examiners with more training or experience to make fewer inconclusive calls. The distributions of number of inconclusive calls were not substantially different between examiners who identified themselves as AFTE members (183) and those that said they were not (7), examiners who identified themselves as working in ASCLD participating agencies (183) and those that said they were not (29), and examiners that were identified as AFTE certified (48) and those that were not (164). Not every respondent answered every survey question.

Table II. Data Collection Information

Enrolled participants	284
Responding participants	218
Known same-source responses (5x218)	1090
Known different-source responses (10x218 – 2 blanks)	2178
Suitability of knowns responses (3x15x218 – 108 blanks)	9702
False positives	22
False negatives	4
Inconclusive responses for known same-source comparisons	11

Table III. Rates and Confidence Intervals

	Rate	95% Confidence Interval*
False Negatives	0.3670%	0.1001 to 0.9369%
False Negatives + inconclusive	1.376%	0.7722 to 2.260%
Poor mark production	2.319%	2.174 to 2.827%
False Positives	0.939% (1.010%)**	0.360% to 2.261%

* Confidence intervals for false negatives, false negatives plus inconclusive results for known same source samples, and poor mark production are calculated using the widely accepted Clopper-Pearson method for calculation of confidence intervals [7].

** As discussed in the text, we report a maximum likelihood estimator derived from a beta-binomial model with the raw proportion of errors in parentheses.

Calculations were made using the R statistical programming system (http://R-project.org). Clopper-Pearson confidence limits for the false-negative probability were computed using the binom.test routine; profile likelihood was used to compute the maximum likelihood estimate and confidence interval for the false-positive probability using the betabinom routine, which is in the VGAM library.

The distribution of unsuitable knowns was randomly distributed in the sample sets provided and did not affect either the false-negative or false-positive rate disproportionately. Of the 1090 known same-source sets analyzed, 66 (6.1%) included less than 3 suitable knowns. Of the 2178 known different-source sets analyzed, 148 (6.8%) included less than 3 suitable knowns. These are not significant differences. No correlation was observed between the occurrence of errors and the rate of poorly marked knowns by a particular firearm or group of firearms. Only two of the 26 errors (false-negative and -positive combined) corresponded with reports of less than 3 suitable knowns (in both cases the examiners reported 2 useable knowns).

<u>Conclusions:</u>

One of the goals of this project was determination of a false-negative rate for cartridge case comparisons. While we have a result based on the frequency of reporting of a finding of elimination for cartridge cases fired from the same handgun, this result remains in some doubt since this rate, even when combined with inconclusive findings from same-source comparisons, is smaller than the measured rate of production of fired cartridge cases unsuitable for comparison. Even if this poor pattern reproduction is not the cause of all of the reported false negatives, it does suggest that the rate we have measured is at best an upper limit on a true false-negative rate purely due to examiner error. One could design a study where the test materials are verified to be highly useable with very clear markings. However, the value of such a measurement is questionable. Since firearms in general are subject to this variability in mark production, if this phenomenon is coupled to the rate of reported false negatives, it may be significantly more important to study how variable this rate is within a particular model and between different models of firearms, instead of eliminating it from future study designs in order to measure a phenomenon (pure examiner error in finding a matching pattern) which is significantly less prevalent. Future studies with different firearms could be carried out with three or more of each model of several popular or frequently encountered firearms. The total number of examiners involved and amount of ammunition used might be much smaller than that used in this study and still be significant. A related study could also be designed to determine how different ammunition might affect the results for a subset of the study firearms. This might include cartridge cases in steel and brass, different manufacturers, and different batches in order to determine the significance of these variables.

Another goal of the project was to measure a false-positive rate for examiner cartridge case comparisons. This rate was measured and for the pool of participants used in this study the fraction of false positives was approximately 1%. The study was specifically designed to allow us to measure not simply a single number from a large number of comparisons, but also to provide statistical insight into the distribution and variability in false-positive error rates. The result is that we can tell that the overall fraction is not necessarily representative of a rate for each examiner in the pool. Instead, examination of the data shows that the rate is a highly heterogeneous mixture of a few examiners with higher rates and most examiners with much lower error rates. This finding does <u>not</u> mean that 1% of the time each examiner will make a false-positive error. Nor does it mean that 1% of the time laboratories or agencies would report false positives, since this study did not include standard or existing quality assurance procedures, such as peer review or blind reanalysis. What this result does suggest is that quality assurance is extremely important in firearms analysis and that an effective QA system must include the means to identify and correct issues with sufficient monitoring, proficiency testing, and checking in order to find false-positive errors that may be occurring at or below the rates observed in this study. Future research could be designed to determine the effectiveness of such quality assurance measures in finding and correcting these issues. If they involve peer review, these studies should examine potential effects like confirmation bias.

The final significant outcome of this study is the variability exhibited in the use of the inconclusive result among actively practicing firearms examiners. While it is entirely possible that examiners treated this study differently than they would casework, there are indications that the guidance of AFTE and SWGGUN is not uniformly followed or implemented by individuals or agencies. The rationale for the use of inconclusive is clear and justifiable as stated by AFTE. Some examiners or agencies choose not to use it at all and declare a comparison without sufficient detail for attribution to a common source to be an elimination. Others never used elimination during this study. Either of these alternatives may have been required by agency policy for these groups of participants. A large group of participants used both inconclusive and elimination as conclusions, indicating varying levels of certainty in the examinations. The wide variation in the use of these conclusions suggests a variation in interpretation that could weaken the basis of the range of conclusions as a standard. This issue is something that the discipline should address in order to clarify and strengthen the basis of all examiners' findings.

For the same-source samples, all four false-negative errors (false eliminations) were made by examiners who did not use inconclusive for any of their responses. This at least suggests that their agencies require that they declare that the comparison result in an identification or an elimination with no middle ground for poorly marked questioned samples, for which there is insufficient detail present to make such a judgment. Given the rate of unsuitable knowns reported for these same firearms, it is quite possible that these agencies are directing these examiners to declare an opinion about the source of casings based on samples with insufficient detail to make a definitive comparison. This can be likened to a fingerprint examiner being asked to identify an individual as either the same person or definitely not the same person who left a partial print at a scene of a crime when that print may be completely smudged or so fragmentary, with so little detail, that other reasonable examiners would say the print is unusable. This would not be a sound implementation of scientific analysis. The 11 examiners who reported inconclusive conclusions for the same-source samples have provided a result that is more descriptive of the evidence presented for poorly marked questioned samples. Although implementation of restrictions on conclusions is a matter of policy rather than of statistical analysis, it is entirely possible that no false-negative responses would have been received in this study had the range of conclusions used by all participants been reflective of the observations of the trained examiners involved instead of current policies. On the other hand, for the different-source samples, 735 were reported as inconclusive and 1421 were reported as eliminations. The fraction of samples reported as inconclusive cannot be attributed to a large fraction of poorly marked knowns or questioned samples in this group. Instead this fraction reflects the large number of examiners who through training or agency policy choose only to report inconclusive conclusions in the absence of an identification or class characteristics differences that would lead to an elimination. We suggest that it may be worth considering whether the use of inconclusive as a matter of policy on such a wide range of observations makes its use less effective in communicating exactly what is inconclusive about the analysis. Some examiners at a national AFTE meeting, when asked by one of the authors of this report about why they would use a mixture of inconclusive and elimination, suggested that policies that restrict conclusions to just identification or inconclusive are too restrictive and don't provide enough detail to report actual observations. Although this is once again a matter of policy rather than statistical significance, we suggest that sound scientific reporting would require conclusions describe observations and that policy makers would be better served by accurate analyses than by limited reporting. In the interests of standardization and clear communication, the means of expressing the findings may need to be uniform, but the present circumstances may be leading to misunderstandings about the meaning of inconclusive through its overuse.

The sample set created for this study, known as the DFSC Ruger SR9 Cartridge Case Set, is a valuable asset that will be used in future research. Not only is it a sizable collection of cartridge cases for which we know the source information regarding which firearm was used to generate each sample, but we have information from professional firearms examiners on the quality and detail of marks and comparisons with same-source and different-source samples. For individuals interested in using the sample set for research please email Henry.P.Maynard2.ctr@mail.mil or usarmy.gillem.dfsc.mbx.email@mail.mil. One of the first follow-up projects will involve collection of high quality 3D digital images of the sample to aid in the development of algorithms for computer comparisons and quantification of the amount of detail available for comparisons. This information could be used for evaluation of human response thresholds and the development of software tools to aid in comparisons.

References:

1) J. Gouwe, J. Hamby, and S. Norris, “Comparison of 10,000 Consecutively Fired Cartridge Cases from a Model 22 Glock .40 S&W Caliber Semiautomatic Pistol,” *AFTE Journal* **40** (1), 57-63 (2008).

2) Bunch, S. G., Murphy D., “A Comprehensive Validity Study for the Forensic Examination of Cartridge Cases,” *AFTE Journal* **35** (2), 201-203 (2003).

3) James Hamby and James Thorpe, "The Examination, Evaluation and Identification of 9mm Cartridge Cases Fired from 617 Different Glock Model 17 & 19 Semiautomatic Pistols", *AFTE Journal* **41** (4) 310-324 (2009).

4) LaPorte, D., “An Empirical Validation Study of Breechface Marks on .380 ACP Caliber Cartridge Cases Fired from Ten Consecutively Finished Hi-Point Model C9 Pistols,” *AFTE Journal* **43** (4), (2011).

5) “An Empirical Study to Improve the Scientific Foundation of Forensic Firearm and Tool Mark Identification Utilizing 10 Consecutively Manufactured Slides”, Award Number 2009-DN-BX-K230, Final Report, Submitted By: Miami-Dade Police Department Crime Laboratory, 2011. https://www.ncjrs.gov/pdffiles1/nij/grants/237960.pdf

6) Committee on Identifying the Needs of the Forensic Sciences Community, National Research Council, “Strengthening Forensic Science in the United States: A Path Forward,” National Academies Press, p. 154 (2009).

7) Clopper, C. J., and Pearson, E. S., “The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial,” *Biometrika*, 26, 404-413 (1934).

Appendix A

Invitation and Informed Consent

THE Ames Laboratory
Creating Materials & Energy Solutions
U.S. DEPARTMENT OF ENERGY

June 14, 2013

Dear Firearm Examiner,

You are being invited to participate in a reliability study evaluating false positive and false negative errors in comparing cartridge casings. Only firearm examiners who currently conduct examinations and are members of AFTE or are employed in the firearms section of an ASCLD member crime laboratory are being asked to participate. This study is sponsored by the Defense Forensic Office and will be carried out by the Midwest Forensics Resource Center (MFRC) at the Ames Laboratory USDOE.

Participation is completely voluntary. There is no compensation for participating in this study. Participating examiners will be sent a set of cartridge casings and asked to compare known and questioned casings. Through a survey instrument, you will be asked to conclude whether the compared casings are identifications, inconclusive, eliminations, or unsuitable. Reported results and findings will be completely anonymized. Additional information about experience, certification, lab accreditation, method, and instrumentation will also be collected through the survey instrument. The study findings will result in a peer-reviewed publication that will be relevant to the legal admissibility of such analyzes.

If you are interested in participating in this study, please complete the attached consent form and return to the MFRC at mfrc@ameslab.gov or by mail (Attn: Firearms Study, 127 Spedding Hall, Ames, IA 50011-3020) by August 31, 2013. If you need further information, please contact Dr. David Baldwin – MFRC Director (dbaldwin@ameslab.gov; 515-294-2069).

Sincerely,

David P. Baldwin

IOWA STATE UNIVERSITY
OF SCIENCE AND TECHNOLOGY

Ames Laboratory is operated by Iowa State University for the U.S. Department of Energy. Ames, Iowa 50011-3020

CONSENT FORM FOR: Investigation of False Positive and False Negative Errors for Cartridge Casing Comparisons

This form describes a research project. It has information to help you decide whether or not you wish to participate. Research studies include only people who choose to take part—your participation is completely voluntary. Please discuss any questions you have about the study or about this form with the project staff before deciding to participate.

Who is conducting this study?

This study is being conducted by Dr. David P. Baldwin from the Midwest Forensics Resource Center (MFRC) at the Ames Laboratory USDOE. This study is funded by the Defense Forensic Office.

Why am I invited to participate in this study?

You are being asked to take part in this study because you are either an Association of Firearm and Tool Mark Examiners (AFTE) member or perform firearm examinations at an ASCLD member crime laboratory. You should not participate if you are not currently performing firearm examinations as part of your normal employment duties.

What is the purpose of this study?

The purpose of this study is to evaluate the reliability of firearms examiners in the analysis and comparison of cartridge casings in order to determine error rates and the degree of correlation between these rates and various related factors.

What will I be asked to do?

If you agree to participate, you will be sent a set of cartridge casings and asked to compare known and questioned casings. This set will contain fifteen (15) sets of cartridge casings for comparison. Through a survey instrument, you will be asked to conclude whether the compared casings are identifications, inconclusive, eliminations, or unsuitable. For microscopic examination, reported results and findings will be completely anonymized. Additional information about experience, certification, lab accreditation, method, and instrumentation will also be collected through the survey instrument.

Your participation will last for the length of time it takes to examine the sets of casings provided and to fill out an answer sheet and survey.

What are the possible risks and benefits of my participation?

Risks—there are no known risks related to your participation in this research.

Benefits—you may not receive any direct benefit from taking part in this study. We hope that this research will benefit society by providing a better statistical evaluation of this common and important forensic discipline that will strengthen the legal system in its understanding of the value of firearms comparisons.

How will the information I provide be used?

The information you provide will be used for the following purposes: Perform a statistical analysis in order to determine error rates and degree of correlation with these rates with various related factors.

What measures will be taken to ensure the confidentiality of the data or to protect my privacy?

Records identifying participants will be kept confidential to the extent allowed by applicable laws and regulations. Records will not be made publicly available. However, federal government regulatory agencies [DOE, DoD], auditing departments of Iowa State University, and the ISU Institutional Review Board (a committee that reviews and approves research studies with human subjects) may inspect and/or copy your records for quality assurance and analysis. These records may contain private information. However, at no point will your survey responses or the results of your examinations be connected to your identity. Only the fact that you participated may be revealed by these reviews and audits. Your participation is only known via this signed consent form. Your identity will remain confidential if these results are published.

To ensure confidentiality to the extent allowed by law, the following measures will be taken: 1) Participant contact information will be kept on a password-protected computer and only accessed by administrative staff; 2) Cartridge casing sample set identifiers will not be linked to any participant; 3) Survey instruments will not contain any identifiable information and will be stored in a locked filing cabinet and access limited to the researchers of this study. This information will be kept for three years after completion of the project. If the results are published, your identity will remain confidential.

Will I incur any costs from participating or will I be compensated?

You will not have any costs from participating in this study. You will not be compensated for participating in this study.

What are my rights as a human research participant?

Participating in this study is completely voluntary. You may choose not to take part in the study or to stop participating at any time, for any reason, without penalty or negative consequences. You can skip any questions that you do not wish to answer.

Whom can I call if I have questions or problems?

You are encouraged to ask questions at any time during this study.

- For further information about the study contact Dr. David Baldwin – MFRC Director (dbaldwin@ameslab.gov; 515-294-2069).
- If you have any questions about the rights of research subjects or research-related injury, please contact the IRB Administrator, (515) 294-4566, IRB@iastate.edu, or Director, (515) 294-3115, Office for Responsible Research, 1138 Pearson Hall, Iowa State University, Ames, Iowa 50011.

Consent and Authorization Provisions

Your signature indicates that you voluntarily agree to participate in this study, that the study has been explained to you, that you have been given the time to read the document and that your questions have been satisfactorily answered. You will receive a copy of the written informed consent prior to your participation in the study.

Additionally, your signature indicates that upon completion of your participation in this study, you agree not to discuss with other examiners that may be participating details about this study or your findings, so that their contribution and findings may be unbiased and independent.

Participant's Name (printed) ____

Participant's Phone No.________________ Participant's email_____________________

______________________________________ ______________________

(Participant's Signature) (Date)

______________________________________ ______________________

(Signature of Lab Director or Section Supervisor *required only when applicable or necessary*) (Date)

Shipping Address to receive study materials (UPS – No P.O. Box addresses, please)

____

____

____

Please return signed form to mfrc@ameslab.gov or mail to:

MFRC, Attn: Firearms
127 Spedding Hall
Ames, IA, 50011-3020

Appendix B

Table of Firearm Serial Numbers and Ammunition Batch Information

Gun	Letter Designation	Serial Number	Ammo Lot#		
1	A1	331-96383	3503 ODFL	2195980	
2	A2	331-96385	3503 ODFL	2195980	
3	A3	331-96387	3503 ODFL	2195980	
4	A4	331-96388	3503 ODFL	2195980	
5	A5	331-96584	3503 ODFL	2195980	
6	B1	331-96585	3503 ODFL	2195980	
7	B2	331-96586	3503 ODFL	2195980	
8	B3	331-96590	3503 ODFL	2195980	
9	B4	331-96592	3503 ODFL	2195980	(1st 400)
			350 ODFL	32206416	(2nd 400)
10	B5	331-96593	350 ODFL	32206416	
11	C1	331-96594	350 ODFL	32206416	
12	C2	331-96604	350 ODFL	32206416	
13	C3	331-96620	350 ODFL	32206416	
14	C4	331-96649	350 ODFL	32206416	
15	C5	331-96651	350 ODFL	32206416	
16	D1	331-96661	350 ODFL	32206416	
17	D2	331-96663	350 ODFL	32206416	
18	D3	331-96664	350 ODFL	32206416	
19	D4	331-96665	350 ODFL	32206416	
20	D5	331-96667	350 ODFL	32206416	
21	E1	331-96669	350 ODFL	32206416	
22	E2	331-96681	350 ODFL	32206416	
23	E3	331-96689	350 ODFL	32206416	
24	E4	331-96718	350 ODFL	32206416	
25	E5	331-96719	350 ODFL	32206416	

Appendix C

Survey and Answer Form

Comparison Group No. ______________________

SURVEY QUESTIONS:

Years Experience:______ Years Training:______

AFTE Certified: Yes☐ No☐ ABC Certified: Yes☐ No☐ Other: ________________

Attended the FBI Specialized Techniques School: Yes☐ No☐ CMS Trained: Yes☐ No☐

Do you work in a firearms ASCLD-member laboratory: Yes☐ No☐

Do you currently conduct firearms casework: Yes☐ No☐

Do you examine other types of evidence: Yes☐ No☐ If Yes, what other types ____________

____

Brand & Model of Microscope used:____________________________________

Type of Lighting used:____

Did you use Pattern Matching, CMS or Both for this test? ____________________

Work performed in accredited laboratory: Yes☐ No☐ Are you an AFTE member: Yes☐ No☐

Set No. Number of known's with sufficient reproduced detail for comparison: 0☐ 1☐ 2☐ 3☐

1 Identification ☐ Elimination ☐ Inconclusive (Please provide basis) Unsuitable ☐

a) ☐ Some agreement of individual characteristics and all discernible class characteristics, but insufficient for an identification.
b) ☐ Agreement of all discernible class characteristics without agreement or disagreement of individual characteristics due to an absence, insufficiency, or lack of reproducibility.
c) ☐ Agreement of all discernible class characteristics and disagreement of individual characteristics, but insufficient for an elimination.

Set No. Number of known's with sufficient reproduced detail for comparison: 0☐ 1☐ 2☐ 3☐

2 Identification ☐ Elimination ☐ Inconclusive (Please provide basis) Unsuitable ☐

a) ☐ Some agreement of individual characteristics and all discernible class characteristics, but insufficient for an identification.
b) ☐ Agreement of all discernible class characteristics without agreement or disagreement of individual characteristics due to an absence, insufficiency, or lack of reproducibility.
c) ☐ Agreement of all discernible class characteristics and disagreement of individual characteristics, but insufficient for an elimination.

Set No. Number of known's with sufficient reproduced detail for comparison: 0☐ 1☐ 2☐ 3☐

3 Identification ☐ Elimination ☐ Inconclusive (Please provide basis) Unsuitable ☐

a) ☐ Some agreement of individual characteristics and all discernible class characteristics, but insufficient for an identification.
b) ☐ Agreement of all discernible class characteristics without agreement or disagreement of individual characteristics due to an absence, insufficiency, or lack of reproducibility.
c) ☐ Agreement of all discernible class characteristics and disagreement of individual characteristics, but insufficient for an elimination.

(over)

Set No.	Number of known's with sufficient reproduced detail for comparison: 0 ☐ 1☐ 2☐ 3☐			
4	Identification	Elimination	Inconclusive (Please provide basis)	Unsuitable
	☐	☐	a) ☐ Some agreement of individual characteristics and all discernible class characteristics, but insufficient for an identification. b) ☐ Agreement of all discernible class characteristics without agreement or disagreement of individual characteristics due to an absence, insufficiency, or lack of reproducibility. c) ☐ Agreement of all discernible class characteristics and disagreement of individual characteristics, but insufficient for an elimination.	☐

Set No.	Number of known's with sufficient reproduced detail for comparison: 0 ☐ 1☐ 2☐ 3☐			
5	Identification	Elimination	Inconclusive (Please provide basis)	Unsuitable
	☐	☐	a) ☐ Some agreement of individual characteristics and all discernible class characteristics, but insufficient for an identification. b) ☐ Agreement of all discernible class characteristics without agreement or disagreement of individual characteristics due to an absence, insufficiency, or lack of reproducibility. c) ☐ Agreement of all discernible class characteristics and disagreement of individual characteristics, but insufficient for an elimination.	☐

Set No.	Number of known's with sufficient reproduced detail for comparison: 0 ☐ 1☐ 2☐ 3☐			
6	Identification	Elimination	Inconclusive (Please provide basis)	Unsuitable
	☐	☐	a) ☐ Some agreement of individual characteristics and all discernible class characteristics, but insufficient for an identification. b) ☐ Agreement of all discernible class characteristics without agreement or disagreement of individual characteristics due to an absence, insufficiency, or lack of reproducibility. c) ☐ Agreement of all discernible class characteristics and disagreement of individual characteristics, but insufficient for an elimination.	☐

Set No.	Number of known's with sufficient reproduced detail for comparison: 0 ☐ 1☐ 2☐ 3☐			
7	Identification	Elimination	Inconclusive (Please provide basis)	Unsuitable
	☐	☐	a) ☐ Some agreement of individual characteristics and all discernible class characteristics, but insufficient for an identification. b) ☐ Agreement of all discernible class characteristics without agreement or disagreement of individual characteristics due to an absence, insufficiency, or lack of reproducibility. c) ☐ Agreement of all discernible class characteristics and disagreement of individual characteristics, but insufficient for an elimination.	☐

Set No.	Number of known's with sufficient reproduced detail for comparison: 0 ☐ 1☐ 2☐ 3☐			
8	Identification	Elimination	Inconclusive (Please provide basis)	
	☐	☐	a) ☐ Some agreement of individual characteristics and all discernible class characteristics, but insufficient for an identification. b) ☐ Agreement of all discernible class characteristics without agreement or disagreement of individual characteristics due to an absence, insufficiency, or lack of reproducibility. c) ☐ Agreement of all discernible class characteristics and disagreement of individual characteristics, but insufficient for an elimination.	☐

Set No.	Number of known's with sufficient reproduced detail for comparison: 0 ☐ 1☐ 2☐ 3☐			
9	Identification	Elimination	Inconclusive (Please provide basis)	Unsuitable
	☐	☐	a) ☐ Some agreement of individual characteristics and all discernible class characteristics, but insufficient for an identification.	☐

(next page)

b) ☐ Agreement of all discernible class characteristics without agreement or disagreement of individual characteristics due to an absence, insufficiency, or lack of reproducibility.
c) ☐ Agreement of all discernible class characteristics and disagreement of individual characteristics, but insufficient for an elimination.

Set No. Number of known's with sufficient reproduced detail for comparison: 0 ☐ 1☐ 2☐ 3☐

10	Identification	Elimination	Inconclusive (Please provide basis)	Unsuitable
	☐	☐	a) ☐ Some agreement of individual characteristics and all discernible class characteristics, but insufficient for an identification. b) ☐ Agreement of all discernible class characteristics without agreement or disagreement of individual characteristics due to an absence, insufficiency, or lack of reproducibility. c) ☐ Agreement of all discernible class characteristics and disagreement of individual characteristics, but insufficient for an elimination.	☐

Set No. Number of known's with sufficient reproduced detail for comparison: 0 ☐ 1☐ 2☐ 3☐

11	Identification	Elimination	Inconclusive (Please provide basis)	Unsuitable
	☐	☐	a) ☐ Some agreement of individual characteristics and all discernible class characteristics, but insufficient for an identification. b) ☐ Agreement of all discernible class characteristics without agreement or disagreement of individual characteristics due to an absence, insufficiency, or lack of reproducibility. c) ☐ Agreement of all discernible class characteristics and disagreement of individual characteristics, but insufficient for an elimination.	☐

Set No. Number of known's with sufficient reproduced detail for comparison: 0 ☐ 1☐ 2☐ 3☐

12	Identification	Elimination	Inconclusive (Please provide basis)	Unsuitable
	☐	☐	a) ☐ Some agreement of individual characteristics and all discernible class characteristics, but insufficient for an identification. b) ☐ Agreement of all discernible class characteristics without agreement or disagreement of individual characteristics due to an absence, insufficiency, or lack of reproducibility. c) ☐ Agreement of all discernible class characteristics and disagreement of individual characteristics, but insufficient for an elimination.	☐

Set No. Number of known's with sufficient reproduced detail for comparison: 0 ☐ 1☐ 2☐ 3☐

13	Identification	Elimination	Inconclusive (Please provide basis)	Unsuitable
	☐	☐	a) ☐ Some agreement of individual characteristics and all discernible class characteristics, but insufficient for an identification. b) ☐ Agreement of all discernible class characteristics without agreement or disagreement of individual characteristics due to an absence, insufficiency, or lack of reproducibility. c) ☐ Agreement of all discernible class characteristics and disagreement of individual characteristics, but insufficient for an elimination.	☐

Set No. Number of known's with sufficient reproduced detail for comparison: 0 ☐ 1☐ 2☐ 3☐

14	Identification	Elimination	Inconclusive (Please provide basis)	Unsuitable
	☐	☐	a) ☐ Some agreement of individual characteristics and all discernible class characteristics, but insufficient for an identification. b) ☐ Agreement of all discernible class characteristics without agreement or disagreement of individual characteristics due to an absence, insufficiency, or lack of reproducibility. c) ☐ Agreement of all discernible class characteristics and disagreement	☐

(over)

of individual characteristics, but insufficient for an elimination.

Set No.	Number of known's with sufficient reproduced detail for comparison: 0☐ 1☐ 2☐ 3☐			
15	Identification	Elimination	Inconclusive (Please provide basis)	Unsuitable
	☐	☐	a) ☐ Some agreement of individual characteristics and all discernible class characteristics, but insufficient for an identification. b) ☐ Agreement of all discernible class characteristics without agreement or disagreement of individual characteristics due to an absence, insufficiency, or lack of reproducibility. c) ☐ Agreement of all discernible class characteristics and disagreement of individual characteristics, but insufficient for an elimination.	☐

(next page)

Appendix D

Participant Instructions

<u>INSTRUCTION SHEET</u>

ITEMS CONTAINED IN THIS PACKAGE:

-This instruction sheet.
-A sealed and unmarked envelope containing the cartridge cases for comparison, a survey, and a recording sheet to record your findings on.
-An unsealed and labeled UPS Express shipping box and an unsealed and unmarked envelop <u>to return</u> the cartridge cases, survey, and recording sheet.

HOW TO CONDUCT THE COMPARISONS:

The cartridge cases are divided into sets of three known's (k's –produced from the same firearm) and a questioned (q). Each *set* comes in its own numbered envelope. <u>This number corresponds to the set number on the recording sheet.</u> Each *set* envelop contains two envelopes which contain the k's and q respectively.

Work with one set at a time. There are a total of 15 sets. Return cartridge cases to their respective envelopes/bags after comparison is made in order to minimize mixing up the casings. Record findings for each set on provided sheet.

You are being asked to compare the k's to the q and render a finding of either "Identification, Elimination, Inconclusive, or Unsuitable," as defined in the AFTE Glossary (Appendix 1) "Range of Conclusions Possible when Comparing Toolmarks."

Indicate on the recording sheet the number of k's (none, one, two, or three) that have adequately reproduced marks for comparison and are adequate to confirm an identification between the k's from the known same source.

Finally, if your finding for a particular set is Inconclusive, please select a basis for this finding. The three choices come from the AFTE Glossary (Appendix 1) "Range of Conclusions Possible when Comparing Toolmarks."

NOTICE:

The purpose of this study is to measure error rates for individual examiners. Please do not peer-review or confirm results of your examination.

AFTER COMPARISONS ARE MADE:

Complete survey. Place survey, recording sheet, and cartridge case set in the provided unmarked envelope and seal. Please do not include any indentifying marks that would indicate your identity on either the survey or envelope.

Place sealed envelope (containing materials) into UPS Express shipping box and seal the box. Return materials to MFRC.

Please do *<u>not</u>* discuss your findings or this study with other examiners who may be participating, so that their contribution and findings may be unbiased and independent.

++

If you have any questions, please contact either Ms. Melinda Schlosser (515-296-6372; mschlosser@ameslab.gov) or Dr. Rudi Luyendijk (515-294-2931; rluyendi@ameslab.gov) for assistance. Thank you for your participation!