

Organization of Scientific Area Committees (OSAC) Firearms and Toolmarks Subcommittee

Response to the President's Council of Advisors on Science and Technology (PCAST) Call for Additional References Regarding its Report "Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods"

14 December 2016

The Organization of Scientific Area Committees (OSAC)¹ Firearms and Toolmarks Subcommittee is composed of sixteen forensic practitioners with a combined 307 years of forensic science experience. The practitioners are drawn from federal, state, county, local and private laboratories from across the country. Additionally, the subcommittee includes four non-practitioners with backgrounds in metrology, statistics, and computer science. The subcommittee's composition meets OSAC's goals of diversity of both forensic practitioners and non-practitioners. Given the responsibility of the subcommittee for informing the process of developing standards and guidelines for the forensic discipline of firearm and toolmark identification, we feel it necessary to respond to the report published by the President's Council of Advisors on Science and Technology (PCAST) and the subsequent Request for Information (RFI) distributed by PCAST co-chair Dr. Eric Lander on December 2, 2016.

The PCAST report addresses numerous subjects and seven disciplines of forensic science. We will limit our response to those portions addressing firearm and toolmark identification. We disagree with PCAST's conclusion that "...firearms analysis currently falls short of the criteria for foundational validity, because there is only a single appropriately designed study to measure validity and estimate reliability." This response will outline why we find PCAST's analysis to be inaccurate.

¹ The purpose of the Organization of Scientific Area Committees (OSAC) is "...to strengthen the nation's use of forensic science by providing technical leadership necessary to facilitate the development and promulgation of consensus-based documentary standards and guidelines for forensic science, promoting standards and guidelines that are fit-for-purpose and based on sound scientific principles, promoting the use of OSAC standards and guidelines by accreditation and certification bodies, and establishing and maintaining working relationships with other similar organizations." <https://www.nist.gov/topics/forensic-science/about-osac>

1 Black-Box (Validation) Study Analysis

PCAST analyzed nine firearm black-box studies and concluded that firearms identification “falls short of the criteria for foundational validity.”² We disagree with their position because it ignores critical details within each study and their review falls short in understanding the research value these studies provide when considered in totality. Additionally, other validation studies have been performed that were not addressed by PCAST.^{3,4,5,6,7,8}

1.1 Introduction

Black-box studies (a common type of validation study) use ground truth to evaluate the soundness and accuracy of examinations. PCAST required that a validation study be of “black-box” design and that samples be examined completely independently of each other. PCAST set the following criteria for determining if a forensic science discipline is scientifically valid: 1) at least two black-box studies that allow for the calculation of a False Positive Error Rate (FPR) and 2) an error rate less than 5%⁹. There is no reference or justification to support that this is a generally-accepted standard.

The studies examined by PCAST were categorized into four different types: “within-set,” “set-to-set,” “partly open set,” and “independent/open.” Within these categories, PCAST examined nine validation studies and discounted the data from eight due to test design. PCAST also made errors when summarizing these studies. They did not accurately count the number of responses, or left data out, from four of the nine validation studies used for their analysis. A summary of the errors can be found in Appendix A.

² PCAST Report “Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods,” (September, 2016) Finding 6, pg 112.

³ Lyons, D. J. “The Identification of Consecutively Manufactured Extractors.” *AFTE Journal*, Vol. 41, No. 3 (2009): 246-256.

⁴ Bunch, S. G., and D. Murphy. “A Comprehensive Validity Study for the Forensic Examination of Cartridge Cases.” *AFTE Journal*, Vol. 35, No. 2 (2003): 201-203.

⁵ Mayland, B. and C. Tucker. “Validation of Obturation Marks in Consecutively Reamed Chambers.” *AFTE Journal*, Vol. 44, No. 2 (2012): 167-169.

⁶ Fadul, T. G. “An Empirical Study to Evaluate the Repeatability and Uniqueness of Striations/Impressions Imparted on Consecutively Manufactured Glock EBIS Gun Barrels.” *AFTE Journal*, Vol. 43, No 1 (2011): 37-44.

⁷ Cazes, M. and J. Goudeau. “Validation Study Results from Hi-Point Consecutively Manufactured Slides.” *AFTE Journal*, Vol. 45, No. 2 (2013): 175-177.

⁸ A listing and summary of additional supportive research, and validation studies pertaining to non-firearm toolmarks, can be found in the SWGGUN Admissibility Resource Kit (ARK). <https://afte.org/resources/swggun-ark/testability-of-the-scientific-principle>

⁹ PCAST Report “Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods,” (September, 2016) “Finding 6”, pp 112, Appendix A, pg 152.

Below we summarize PCAST's analysis and why we disagree with their findings.

1.2 Within-set Studies

PCAST summarized two "within-set" validation studies.^{10,11} The PCAST committee could not calculate a False Positive Error Rate (FPR) using these studies, so they did not use them to measure the validity of firearm and toolmark identification.

The dismissal of these studies does not accurately reflect the scientific value of the research. A total of 1037 different-source comparisons were performed. No false identifications or false eliminations were reported by any of the participants. Therefore, these two studies provide empirical and independent support that the overall error rate for firearm and toolmark identification is low, despite the inability to calculate a false positive error rate.

1.3 Set-to-Set Comparison/Closed Set Studies

PCAST summarized four "closed set" studies.^{12,13,14,15} PCAST is critical of these test designs because each comparison is not independent of the others. The assumption is that examiners may be able to deconstruct the test design, and PCAST likens this to the same logic as solving a "Sudoku" puzzle.¹⁶ The analogy used by PCAST misrepresents the challenge posed by these tests. First, three of the studies (Brundage et al., Hamby et al., Fadul et al.) used consecutively manufactured firearms. Consecutively manufactured firearms have been shown to have the potential for subclass characteristics, which are toolmarks that sometimes carry over, with very

¹⁰ Smith, E. "Cartridge case and bullet comparison validation study with firearms submitted in casework." *AFTE Journal*, Vol. 37, No. 2 (2005): 130-5. There were a total of 16 same-source comparisons and 704 different-source comparisons in this study. 13 of the 16 same-source comparisons were correctly identified and 3 were inconclusive. There were no false identifications or false eliminations reported.

¹¹ DeFrance, C.S., and M.D. Van Arsdale. "Validation study of electrochemical rifling." *AFTE Journal*, Vol. 35, No. 1 (2003): 35-7. There were a total of 45 same-source comparisons and 333 different-source comparisons. 42 of the 45 same-source comparisons were correctly identified and 3 were inconclusive. There were no false identifications or false eliminations.

¹² Stroman, A. "Empirically determined frequency of error in cartridge case examinations using a declared double-blind format." *AFTE Journal*, Vol. 46, No. 2 (2014):157-175.

¹³ Brundage, D.J. "The identification of consecutively rifled gun barrels." *AFTE Journal*, Vol. 30, No. 3 (1998): 438-44.

¹⁴ Fadul, T.G., Hernandez, G.A., Stoiloff, S., and S. Gulati. "An empirical study to improve the scientific foundation of forensic firearm and tool mark identification utilizing 10 consecutively manufactured slides." *AFTE Journal*. Vol. 45, No. 4 (2013): 376-93.

¹⁵ Hamby, J.E., Brundage, D.J., and J.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*, Vol. 41, No. 2 (2009): 99-110.

¹⁶ PCAST Report "Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods" (September, 2016) Section 5.5, pp 106. PCAST was quoting Jeff Salyards, Director of the Defense Forensic Science Center.

little change or variation, from one machined part to the next on the same production line.^{17,18,19} Qualified examiners are able to recognize these marks so as not to use them for conclusions of identification. Though consecutively manufactured firearms are not likely to be encountered in actual casework, the authors used them in an attempt to create a worst-case scenario (i.e. potential best known non-matches). Additionally, each test used *more* questioned samples than knowns (15 questioned samples from 10 consecutively manufactured firearms). Therefore, taking these tests was not as simple as figuring out a few of the correct answers and then deducing the rest. Since these tests used consecutively manufactured samples, it was just as important to know if examiners could correctly identify samples as it was to know if samples were falsely identified. This is the reason at least one true match was provided with each questioned cartridge case.

Another study discounted by PCAST was conducted by Stroman et al. This validation study used cartridge cases that had been fired in Smith & Wesson pistols. While this study did not use consecutively-manufactured samples, the firearms were the same make and model and had documented subclass characteristics on the firearms' ejectors. Again, these are potentially difficult samples and provide the opportunity for false positive errors, yet none were observed.

In each of these four studies, the authors attempted to create tests with potentially challenging samples. Each of these studies provide insight into the overall error rate (see Appendix A for more details about each study). The fact that few false positive errors occur is strong evidence in support of the discipline of firearm and toolmark identification. These studies present evidence that firearm and toolmark examiners can reliably and accurately associate questioned toolmarks to the correct source tool. Though the test design does not fit the model proposed by PCAST, these studies present valuable performance estimates and should not be disregarded. When viewed collectively, these studies are independent of each other and show a low overall error rate among the tested examiners. This provides strong support for the overall validity of firearm and toolmark identification.

¹⁷ Weller, T.J., Zheng, X.A., Thompson, R.M., and F. Tulleners. "Confocal microscopy analysis of breech face marks on fired cartridge cases from 10 consecutively manufactured pistol slides." *Journal of Forensic Sciences*, Vol. 57, No. 4 (2012): 912-17. This study has documented subclass characteristics among the 10 consecutively manufactured pistol slides. An eleventh pistol slide, that was not part of the consecutive batch, no longer has the same subclass toolmarks.

¹⁸ Miller J., Beach G. "Toolmarks: Examining The Possibility of Subclass Characteristics" *AFTE Journal*, Vol 32, No 4: 296-345.

¹⁹ Subclass characteristics are features that may be produced during manufacture that are consistent among items fabricated by the same tool in the same approximate state of wear. These features are not determined prior to manufacture and are more restrictive than class characteristics. *AFTE Glossary*, 6th Edition.

1.4 Partly Open Set

PCAST summarized another validation study and categorized it as “partly open.”²⁰ We would like to highlight the fact that this study also uses consecutively manufactured samples and, as described above, provides examiners with test samples which are most likely to have similar toolmarks since the firearms used to create them were sequentially manufactured with the same tools.

PCAST's statistical analysis of this report focused solely on two unknowns that had no matching known. This analysis is incomplete, and differs from the analysis used by PCAST in the “set-to-set/closed set” and “open set” studies where all “conclusive” responses were used to calculate the False Positive Error Rate.

The authors' reported error rate (0.7%) was low and this study provides an additional independent study establishing that firearm and toolmark examiners can accurately associate questioned toolmarks to the correct source tool.

1.5 Open Set

PCAST summarized another validation study and categorized it as “open.”²¹

Each test taker in this study was instructed to work independently and not collaborate with other test takers. These instructions negate an important quality assurance step used in most accredited forensic laboratories: the peer review process known as verification²². Verification is a reevaluation of a comparison by another qualified examiner to ensure there is sufficient data to support the conclusion. Many laboratories accomplish this by direct reexamination of the evidence, while others use representative photographs of sufficient quality for the verification step. The errors reported in this paper may have been caught if verification were allowed. This suggests the true false positive error rate may be lower than calculated in this study. We would like to highlight that Baldwin et al. discusses this point in their study (emphasis added):

“This finding does not 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**

²⁰ Fadul, T.G., Hernandez, G.A., Stoiloff, S., and S. Gulati. “An empirical study to improve the scientific foundation of forensic firearm and tool mark identification utilizing consecutively manufactured Glock EBIS barrels with the same EBIS pattern.” National Institute of Justice Grant #2010-DN-BX-K269, December 2013.

²¹ Baldwin, D.P., Bajic, S.J., Morris, M., and D. Zamzow. “A study of false-positive and false-negative error rates in cartridge case comparisons.” Ames Laboratory, USDOE, Technical Report #IS-5207 (2014) afte.org/uploads/documents/swggun-false-postive-false-negative-usdoe.pdf.

²² In the other validation studies discussed above, verification was also unlikely because test takers were not to collaborate with other test takers.

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.

“²³

It should be noted that PCAST used the data from the study to recalculate a false positive error rate by using only exclusion conclusions and omitting the inconclusive results. This resulted in a rise in the calculated error rate from 1.01% to 1.5%. The different error rates provide different answers for different questions: The Baldwin et al. error rate estimates how often non-matching cartridge cases are falsely identified, while PCAST's error rate estimates the proportion of definitive (i.e. not inconclusive) results that are incorrect when non-matching cartridge cases are examined.

Baldwin et al. provide a discussion about inconclusive results (emphasis added).²⁴

“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).”

Baldwin et al. found the rate of poor quality mark production to be 2.3% (+/- 1.4%). This rate is double the calculated false positive error rate. This provides support for the use of inconclusive results in the calculation of error rates.

We would like to highlight the fact that the Baldwin study found “all but two of the 22 false identification calls were made by five of 218 examiners.”²⁵ This indicates when errors do occur, they may be committed by the same few examiners. This supports the need for rigorous

²³ Baldwin et al. Pg 18.

²⁴ Baldwin et al. Pg 6

²⁵ Baldwin et al. Pg 16.

training, periodic proficiency testing, continuing education and thorough laboratory quality control measures.

1.6 Smith et al. Study

The final validation study examined by PCAST was the Smith et al. study, in which the authors created a test that mimics casework²⁶. PCAST concluded this study was insufficient to test the validity of firearm identification:

“While interesting, the paper clearly is not a black-box study to assess the reliability of firearms analysis to associate ammunition with a particular gun, and its results cannot be compared to previous studies.”²⁷

PCAST recognizes the study as being new and novel. We disagree with their observation that since the study is not a “black-box” design then the study does not provide support for the validity of firearm identification. In the test design that PCAST requires, test takers examine only one questioned sample at a time, independent of other questioned samples. While we understand this test design allows for easier statistical analysis, one to one comparisons are not an accurate representation of actual casework. A typical examination for a firearm examiner entails opening a package of evidence with dozens of items and attempting to associate or disassociate the items. This study tested that process by forcing examiners to make all of the typical decisions they would make in casework, rather than conducting a series of examinations on isolated pairs of specimens. The test takers were presented with bullets and cartridge cases of various ammunition types, and asked to perform both class and individual characteristic evaluations. They were not given any information about the source of any of the items.

Test takers were faced with a real-world scenario and performed very well. Although not stated in the PCAST footnote referencing this article, the overall error rate for this study was 0.303%.

1.7 Conclusions

PCAST reviewed nine validation studies and through their criteria, elected to discount eight of those studies. Two of those disregarded studies (the “within-set” design) had no false positive results. Five of the disregarded studies had very few false positives (see Appendix A) and the last study (which attempted to replicate casework) found a low overall error rate (0.303%).

²⁶ Smith, T., Smith, G.A., Snipes, J.B. “A Validation Study of The Bullet and Cartridge Case Comparisons Using Samples Representative of Actual Casework.” *Journal of Forensic Sciences*, Vol. 61, No. 4: 939-946

²⁷ PCAST Report “Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods” (September, 2016), footnote #335.

When PCAST set criteria for the validity of a forensic science discipline, they chose an arbitrary threshold of having at least two black box studies. The black-box test design favored by PCAST requires that each questioned sample be examined independently from each other. Examiners are not faced with completely independent examinations when they analyze evidence in a case. It is not realistic, if trying to replicate casework, to have fifteen or twenty individual sets of comparisons, each of which is made independent of each other. The PCAST-proposed design may make sense from a purely statistical standpoint, but does not simulate the practical task of an examiner performing casework. The OSAC subcommittee believes that various types of tests are valuable and can provide meaningful information regarding the potential error rates²⁸.

2.0 Subjective and Objective Methods

PCAST defines objective feature comparison methods as “methods consisting of procedures that are each defined with enough standardized and quantifiable detail that they can be performed by either an automated system or human examiners exercising *little or no judgment*” (emphasis added). PCAST defines subjective methods as “methods including key procedures that involve *significant human judgment*”²⁹ (emphasis added).

In fact, all disciplines, including firearm and toolmark identification, require some human judgment or interpretation of results. Implementation of more objective techniques may make those interpretations easier, but judgment will still be required.

We agree, however, with the goal of continuing to research and implement more objective analytical methods. One of our subcommittee's task groups is writing standards that will assist industry and crime laboratories with the validation and implementation of new technology. Additionally, there is a growing body of research using three-dimensional instrumentation and advanced machine-learning algorithms to compare toolmarks. The research fails to disprove the foundational premise of firearm and toolmark identification: that fired ammunition components can be associated to (or eliminated from) the originating firearm through the comparison of microscopic toolmarks. In fact, the recent research provides strong objective

²⁸ Different test designs estimate different error rates. For example: when examining evidence from an officer involved shooting where each officer admits to firing their firearm: error rates based on data from “set to set/closed-set” studies may be more appropriate while the Smith et. al. study may provide a better estimate for an examination of numerous items with no questioned firearm. All of these studies have the potential to provide a relevant error rate estimates and the “true” error rate may not be the same for each situation.

²⁹ PCAST Report “Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods” (September 2016) Section 4.1, pp. 46-47.

support for this premise. The PCAST committee was provided with 25 citations by our subcommittee documenting this work; however, their report only cites two studies.

3.0 AFTE Theory of Identification is Circular Logic

PCAST states that the AFTE Theory of Identification is circular logic. PCAST's summary of the theory makes it sound circular:

"It declares that an examiner may state that two toolmarks have a "common origin" when their features are in "sufficient agreement." It then defines "sufficient agreement" as occurring when the examiner considers it a "practical impossibility" that the toolmarks have different origins."³⁰

The PCAST Report makes the AFTE Theory sound circular by ignoring the basis for "sufficient agreement." This is based on a misunderstanding of what constitutes "sufficient agreement." They claim it is an arbitrary point at which the examiner considers it a "practical impossibility." PCAST seems to believe that this "practical impossibility" is arbitrarily decided by the examiner, thus making the theory sound circular. This is incorrect. The sufficient agreement threshold is exhibited when the amount of agreement is greater than best known non-matches established by the community and conveyed to each examiner through a lengthy and extensive training program. That is, it is not an arbitrary point. In fact, by definition, no non-matches can ever have more similarity than the sufficient agreement point. When the basis for the ground truth is included, the AFTE Theory is not circular.

4.0 Focus on Training and Experience Rather Than Empirical Demonstration of Accuracy

PCAST quote:

"Many practitioners hold an honest belief that they are able to make accurate judgments about identification based on their training and experience."³¹

In all professions, proper training and experience is critical. Firearm and toolmark identification is like other applied sciences (e.g. medicine, engineering) that require training to become proficient and experience to further refine and maintain that proficiency. There is only so much that textbooks can teach, and structured training (like residency for physicians) is a critical aspect of developing proficiency. It is through rigorous training that examiners develop their criteria for what constitutes an elimination, an identification, or an inconclusive result. They

³⁰ PCAST Report "Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods" (September 2016) Section 4.7, pp. 60.

³¹ PCAST Report "Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods" (September 2016) Section 4.7, pp 60-61.

learn and understand the differences in microscopic agreement between toolmarks created by the same source (a known match) and toolmarks created by different sources (a known non-match) and how that understanding factors into any conclusion of elimination, inconclusive, or identification. Examiners do not memorize all patterns that have been observed, as suggested in the PCAST report.

5.0 Conclusion

The Firearms and Toolmarks Subcommittee of OSAC fundamentally disagrees with the conclusions regarding the firearm and toolmark identification discipline presented in the PCAST report. Four major points have been put forth in this response. First, we disagree with the premise that a structured black-box study is the only useful way to gain insight into both the foundations of firearm and toolmark identification and examiner error rates. Taken collectively, the published studies support the underlying principles of firearm and toolmark examination and the fact that examiner error rates are quite low. PCAST's critique of these studies included several misunderstandings. Second, PCAST's dismissal of methods employing a subjective component discounts the core scientific methods that have been used for hundreds of years. Third, PCAST misunderstands and misquotes the AFTE Theory of Identification. PCAST's summary of the AFTE Theory of Identification leaves out important provisions. Fourth, PCAST minimizes the value of training and experience. The training received by firearm examiners includes both subjective and objective components and is comparable to the domain-specific rigor of other applied scientific fields.

We do not agree that firearm identification "...falls short of the criteria for foundational validity." However, we do agree that a hallmark of any scientific endeavor is ongoing research and technology development. Indeed, our subcommittee, which is tasked with writing standards and providing guidance to the profession, would not exist if it was believed that the field of firearm identification is flawless and requires no improvement. As such, we are hopeful that the path forward from the PCAST report is a renewed commitment to research in the forensic sciences, continued testing of foundational principles, and a more robust collaboration between the academic and forensic practitioner communities.

Appendix A

Errors and Omissions in PCAST Summaries of Firearms and Toolmarks Validation Studies

PCAST incorrectly summarized four of the nine validation studies used in their analysis of firearm and toolmark identification. For clarity, we first repeat some of the terms used by PCAST to illustrate how they (and we) calculated these error rates.

“The results of a given empirical study can be summarized by four values: the number of occurrences in the study of true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN)”³²

PCAST used the following formula to calculate the “maximum likelihood estimate of FPR”: $FP/(FP+TN)$.³³ For those unfamiliar with statistics, we recalculate the FPR for the Baldwin et al. study. There were a total of 2178 different-source comparisons performed: 1421 were declared elimination, 735 were reported as inconclusive, and there were 22 false positives reported. PCAST did not use inconclusive results in their statistical treatment (as we discussed in Section 1.5). Therefore, PCAST’s FPR calculation for the Baldwin et al. study is: $FPR = 22/(1421+22)$. This equals 0.015, or 1.5%. Conversely, recognizing that inconclusive results are appropriate³⁴, Baldwin, et al. included inconclusive results in their calculations, as follows: $FPR = 22/(1421+735+22)$. This equals 0.010, or 1.0%.³⁵

*For the “set-to-set/closed” studies, PCAST used correct identifications in lieu of using true negatives*³⁶. PCAST does not explain or justify why they did this. The error rates reported by PCAST for the “set-to-set/closed” studies found in Table 2 on page 111 of the PCAST report are not false positive error rates and should not be reported as such.

Below we summarize the errors made by PCAST in their assessment of four of the nine studies.

³² PCAST Report “Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods”, Appendix A, pg 152.

³³ PCAST Report “Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods”, Appendix A, pg 152.

³⁴ Baldwin et al., pg 6.

³⁵ Baldwin et al., pg 16.

³⁶ See footnote 327 of PCAST report: “Of the 10,230 answers returned across the three studies, there were there were 10,205 correct assignments, 23 inconclusive examinations and 2 false positives.”

[Brundage Study](#)

The PCAST summary of the Brundage Study is (emphasis added):

In this study, bullets were fired from 10 consecutively manufactured 9 millimeter Ruger P-85 semi-automatic pistol barrels. Each of 30 examiners received a test set containing **20 questioned** bullets to compare to a set of **15 standards**, containing at least one bullet fired from each of the 10 guns. Of the **300 answers returned**, there were no incorrect assignments and one inconclusive examination.

This is not correct. The Brundage study consisted of 15 questioned bullets compared to a set of 10 standards (two test fired bullets from each standard set). This test was sent to 30 examiners and 450 answers returned (30 examiners x 15 questioned bullets) with no false positives and one inconclusive conclusion.

[Hamby Study](#)

The Hamby Study was a continuation of the Brundage study. Hamby et al. used the same firearm and ten consecutively manufactured barrels to produce an additional 240 test sets. The PCAST summary of this study states (emphasis added):

In this study, bullets were fired from 10 consecutively rifled Ruger P-85 barrels. Each of **440 examiners** received a test set consisting of 15 questioned bullets and two known standards from each of the 10 guns. Of the **6600 answers returned, there were 6593** correct assignments, seven inconclusive examinations and no false positives.

This study combined the conclusions from the Brundage study, and additional results collected with both the original Brundage test sets and the 240 new test sets. If we subtract the original 30 responses from the Brundage study, the Hamby et al. article reports an additional 477 examiners having completed the test, for a total of 7155 answers with 7148 correct assignments and 7 inconclusive conclusions.

[Fadul Pistol Slides Study](#)

The PCAST summary of the Fadul Pistol Slides Study:

In this study, bullets were fired from 10 consecutively manufactured semi- automatic 9mm Ruger pistol slides. Each of 217 examiners received a test set consisting of 15 questioned cartridge cases and two known cartridge cases from each of the 10 guns. Of the 3255 answers returned, there were 3239 correct assignments, 14 inconclusive examinations and two false positives.

OSAC Firearms and Toolmarks Subcommittee's Response to the PCAST Call for Additional References

This summary is correct; however, it is incomplete because it only includes Phase 1 of the study. It does not include the second phase of the study, the durability study. Results for Phase 1 and 2 are included in the same report. In Phase 2, an additional 114 examiners participated. The examiners received 5 more questioned cartridge cases (after the firearm had been fired 1000 times) and were asked to compare these cartridge cases to the 10 cartridge cases from the knowns that were previously received. A total of 570 answers were returned with 564 correct assignments, 5 inconclusive and one false positive.

[Fadul EBIS Barrels Study](#)

The PCAST summary of this study states (emphasis added):

The 165 examiners in the study were asked to assign a collection of **15 questioned** samples, fired from **10 pistols**, to a collection of known standards; two of the 15 questioned samples came from a gun for which known standards were not provided.

This is not correct. Each test consisted of two known standards from each of the 8 pistols and 10 questioned samples. One of the known pistols had no matching questioned samples. Additionally, two of the unknowns had no matching known pistol.

Fadul et al. reported an overall error rate of 0.7% (95% lower bound 0.2%, 95% upper bound 1.2%).