B108 A Validation Study of the Accuracy, Repeatability, and Reproducibility of Firearm Comparisons

Keith L. Monson, PhD, Quantico, VA 22135; Erich D. Smith, MS, Federal Bureau of Investigation Laboratory, Quantico, VA 22135; Jennifer L. Stephenson, MSFS, Federal Bureau of Investigation, Quantico, VA 22135; L. Scott Chumbley, PhD*, Iowa State University, Ames, IA 50011; Stanley J. Bajic, PhD, Ames Laboratory, Ames, IA 50011; Max D. Morris, PhD, Ames, IA 50011; Daniel S. Zamzow, MS, Ames Laboratory, Ames, IA 50011*

Learning Overview: After attending this presentation, attendees will better understand the design process and results achieved in a large-scale decision analysis ("black box") study involving a large group of firearms examiners.

Impact on the Forensic Science Community: This presentation will impact the forensic science community by providing empirical measurements of the accuracy, repeatability, and reproducibility of analyses performed by firearms examiners for cartridge case and bullet sample sets.

The study hypothesis was that trained/qualified firearms examiners can accurately determine source conclusions (repeatability) when applying the Association of Firearm and Tool Mark Examiners (AFTE) Theory of Identification, as well as reproduce the same result(s) when later encountering the same comparison.

This study complements those conducted previously, adding important additional features. A previous study conducted on the accuracy of firearms examiners was generally viewed favorably by the President's Council of Advisors on Science and Technology (PCAST), but PCAST advised that additional, similarly designed "black box" investigations were required to establish foundational validity.[1,2]

The present study implemented a fully randomized, open set, and double-blind design involving challenging comparisons of fired bullets and cartridge cases. In order to maintain double-blind conditions, the experimental study was conducted under contract by scientific staff at another organization, which sent randomized specimens to participating examiners and performed statistical analysis. Following a call for participants, volunteer active examiners were provided with 15 comparison sets of two known and one unknown cartridge cases fired from a collection of Beretta® and Jimenez® firearms and 15 comparisons sets of two known and one unknown bullets fired from Beretta® and Ruger® firearms. In order to minimize reproduction of marks, the ammunition selected for testing was Wolf Polyformance® 9mm Luger (9x19mm), with acrylic polymer-coated, steel cartridge cases and lead core, copper-plated, steel-jacketed bullets. The firearms and ammunition selected for this study were purposely chosen due to their propensity to produce challenging and ambiguous test samples creating difficult comparisons for examiners. The firearms, bullets, and cartridge cases used for the study were collected by researchers in the first laboratory and delivered to scientists in the contract laboratory, who then conducted the study and engaged in the generation and distribution of test packets and collection and analysis of the data. A total of 173 qualified examiners took part in the study. The participating examiners were asked to follow the provided instructions rather than adhere to their laboratory policies and were instructed not to discuss their results with anyone else in their laboratory. In order to further maintain the double-blind, "black box" nature of the study, the team associated with communicating with the examiners was not aware of the contents of each comparison set, and the experimental / analysis group was never aware of the examiner's identities.

The total number of comparisons carried out was 20,130, of which 8,640 tested for accuracy, 5,700 tested for repeatability, and 5,790 were tested for reproducibility. Definitive false positive error rate estimates that take examiner heterogeneity into account are 0.66% for bullets and 0.93% for cartridge cases. False negative error rate estimates are 2.87% (bullets) and 1.87% (cartridge cases). These estimates are based on data that include comparisons from barrels produced sequentially in time and those separated in the manufacturing process, rounds fired early in the life of a barrel and after many rounds had been fired, and rounds fired from both high and low cost-point firearms. Individual error rates within each of these categories have also been calculated and vary slightly from the overall average in ways that might be expected (e.g., higher error rates are seen for rounds widely separated in firing order than sequentially, lower cost point firearms have a higher false negative error rate than average, etc.). As in earlier studies, the majority of errors were produced by a relatively small number of examiners. The numbers found in the current study are generally consistent with the results reported in prior studies and, therefore, constitute the foundational validation the PCAST report said was lacking.

Reference(s):

1. D.P. Baldwin et al. *A Study of False-Positive and False-Negative Error Rates in Cartridge Case Comparisons.* (https://apps.dtic.mil/dtic/tr/fulltext/u2/a611807.pdf: Ames Laboratory, USDOE, 2014).
2. PCAST. *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods*. (https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf: President's Council of Advisors on Science and Technology, 2016).

Firearms Examination, Error Rate, Black Box