



### B164 Lessons Learned From Conducting Black-Box Evaluations in Multiple Disciplines

R. Austin Hicklin, PhD\*, Noblis, Reston, VA 20191; JoAnn Buscaglia, PhD\*, FBI Laboratory, Counterterrorism and Forensic Science Research Unit, Quantico, VA 22135

**Learning Overview:** The goal of this presentation is to help attendees understand the issues involved in designing and conducting black-box evaluations of the accuracy and reliability of forensic examiners, across several forensic disciplines.

**Impact on the Forensic Science Community:** This presentation will impact the forensic science community by describing approaches to assess potential areas of strength and weakness in multiple areas of forensic science.

**Hypothesis Statement:** Lessons learned from black-box evaluations in a forensic discipline can identify issues for consideration in the design of studies assessing the accuracy and reliability of examiners in other forensic disciplines.

The 2016 President's Council of Advisors on Science and Technology (PCAST) Report on forensic science stated that black-box evaluations are necessary to establish the validity of forensic examination methods that rely on human judgment and provided some guidance on what criteria PCAST required for such testing to be considered rigorous and suitable for court admissibility.<sup>1</sup> The accuracy, reproducibility, and repeatability of conclusions in a number of forensic science disciplines have not yet been assessed using such black-box studies. The results of such studies would assist in supporting the scientific basis of these forensic disciplines and provide insights into improved operational procedures and training.

The forensic science community is currently engaged in a variety of efforts to evaluate the accuracy and reliability of forensic conclusions in the pattern evidence disciplines. This presentation will discuss the issues involved and lessons learned from designing and conducting black-box evaluations in five forensic disciplines, including completed studies of latent print examination and studies in progress of handwriting examination, footwear examination, bullet examination, and bloodstain pattern analysis. This presentation will discuss what aspects of study design and data collection have commonalities that apply to multiple disciplines and the aspects that are specific to individual disciplines.

This presentation will discuss topics that should be considered in the design of black-box evaluations, including: (1) controlled collection of samples to simulate operational casework; (2) accounting for different conclusion scales; (3) quality assurance; (4) analysis methods; (5) distribution of physical vs. electronic samples; and (6) evaluation based on classification of individual samples (e.g., bloodstain pattern analysis) vs. comparison of multiple samples.

These studies are designed to evaluate and/or demonstrate the validity of each discipline, as well as providing essential information to laboratory managers and practitioners that may be used to improve or enhance practices in the laboratory. This presentation will also discuss how the results of black-box studies provide data for subsequent "white-box" evaluations and present lessons learned from designing and conducting white-box testing.

#### Reference(s):

<sup>1</sup>. Executive Office of the President, President's Council of Advisors on Science and Technology (2016) Report to the President. *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods*. September 2016.

#### Evaluation, Error Rates, Examiner Accuracy