A Statistical Method for Evaluating the Reliability of Identification of Randomly Acquired Characteristics in Footwear Impression Evidence

Corey B. Katz, Naomi Kaplan-Damary and Hal S. Stern

Learning Objective: Participants in this session will learn about a statistical approach to quantifying the reliability of shoe print examinations, specifically the consistency with which randomly acquired characteristics (RACs) are identified by examiners.

Impact Statement: This presentation will impact the forensic science community by offering the statistical foundation of a method for evaluating the performance of footwear examiners in identifying randomly acquired characteristics. This is an important contribution to learning about the reliability of footwear examinations.

There has recently been increased attention on the reliability and validity of methods used in the analysis of forensic evidence. Black-box studies, studies in which questioned-known evidence pairs for which the underlying truth is known are analyzed by forensic examiners, have emerged as a powerful tool to learn about intra-examiner reliability, inter-examiner reliability, and validity. Results of such studies can be complemented by additional studies that focus on performance of examiners on different aspects of the examination. This presentation focuses on the development of a statistical approach to assessing the performance of examiners in identifying randomly acquired characteristics on footwearimpressions.

Footwear examiners are asked to compare a footwear impression found at a crime scene (sometimes known as the questioned impression) with an impression from one or more suspect (known) shoes. The first step in this process is to rule out any shoes that do not match the basic characteristics (size or tread pattern) of the questioned impressions. Examiners next consider wear. A critical part of the examination process is the identification and examination of randomly acquired characteristics (RACs), markings that indicate scratches or holes that have formed on the bottom of shoe soles as the shoe is worn. If the questioned and known impressions have RACs in the same location, then this is evidence that supports the hypothesis that the two impressions have a commonsource.

Identifying RACs takes a great deal of training and, in many cases, is a difficult and time-consuming process. Footwear examiners record the location, type (holes, scratches, etc.), and other key attributes of each RAC. The proficiency tests that exist are designed to evaluate examiners performance when it comes to their ability to correctly match a questioned impression to a putative source. While RACs are considered in this decision-making process, there has been little research done in evaluating the performance and reliability of footwear examiners when it comes to the identification of RACs on a footwear impression. This performance assessment is complicated by the lack of ground truth of RACoccurrence.

The method described in this presentation is a first step in estimating the performance of examiners in identifying RACs examinations. It is based on a probability model that assumes there are repeated annotations of the same impression either by a single examiner or by multiple examiners. Importantly, the approach does not require that there be an a priori definition of ground truth. The probability model builds on an approach developed in the context of brain imaging.[1] Data from a pilot study conducted by the Division of Identification and Forensic Science of the Israel Police (DIFS) allows for the estimation of performance because multiple student trainees identified RACs on the same footwear impression and each examiner examined multiple shoes. Using this data, performance parameters were defined as the ability of the examiner to determine whether or not RACs appear on the impression. Further, the model can incorporate information from multiple footwear impressions. Preliminary results have indicated this approach is a step in the right direction regarding estimation of the performance of footwear examiners.

References

1. Simon K. Warfield, Kelly H. Zou, and William M. Wells. "Simultaneous Truth and Performance Level Estimation (STAPLE): An Algorithm for Validation of Image Segmentation." *IEEE Trans Med Imaging* 23 no. 7 (2004): 903-921.

Keywords: Performance Assessment, Examiner Reliability, Randomly Acquired Characteristics