2023

Estimate of the Random Match Frequency of Acquired Characteristics in a Forensic Footwear Database

Alyssa N. Smale
West Virginia University, ans00016@mix.wvu.edu

Estimate of the Random Match Frequency of Acquired Characteristics in a Forensic Footwear Database

Alyssa N. Smale, M.P.S.

Dissertation submitted
to the Eberly College of Arts and Sciences
at West Virginia University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in
Forensic & Investigative Science

Dr. Jacqueline A. Speir, Ph.D., Chair

Dr. Glen P. Jackson, Ph.D.

Dr. Kenneth J. Ryan, Ph.D.

Dr. Hariharan K. Iyer, Ph.D.

Department of Forensic & Investigative Science

Morgantown, West Virginia
2023

Keywords: footwear; randomly acquired characteristics; random match frequency; non-mated pairs; percent area overlap; simulated crime scene impressions

ABSTRACT

Estimate of the Random Match Frequency of Acquired Characteristics in a Forensic Footwear Database

Alyssa N. Smale, M.P.S.

When analyzing footwear impression evidence, one of the goals of an examiner is to determine if an exemplar shoe could be the source of an impression found at a crime scene. This opinion is based on an assessment of the similarity of class characteristics and randomly acquired characteristics (RACs) between the known and questioned impressions, as well as the rarity of the observed characteristics. The primary aim of this research was to estimate the random match frequency of randomly acquired characteristics (RAC-RMF) within a forensic footwear database to determine the frequency of RACs with geometric similarity occurring in the same relative position on unrelated outsoles, which could potentially increase the chance of an erroneous source association.

RAC-RMF was estimated using high-quality test impressions of the 1,300 shoes in the West Virginia University (WVU) footwear database. Each impression in the database was sequentially held out and compared to the remaining 1,299 impressions to determine if unrelated shoes possessed similar RACs in the same relative locations. With over 80,000 RACs available for analysis, this resulted in nearly four million comparisons, which were performed using a combination of visual comparisons and predictions from a mathematical model based on a percent area overlap similarity score. Nearly 70% of the shoes in the database shared an indistinguishable pair with at least 1 out of 1,299 unrelated shoes, with a maximum RAC-RMF value of 49 out of 1,299 observed, and up to 5 indistinguishable RAC pairs shared between unrelated outsoles.

A similar evaluation was performed on two simulated crime scene impression datasets each containing more than 160 impressions deposited in blood or dust, respectively. A total of 759 RACs were identified in blood impressions created on tile, leading to over 77,000 non-mated RAC comparisons between blood impressions and test impressions from 1,299 unrelated outsoles. RACs in blood impressions were smaller on average than their test impression mates, and therefore exhibited a 66% increase in the number of indistinguishable RAC pairs. Depending on RAC length, relative RAC-RMFs of at least 0.0008 were encountered at a rate between 3.4% and 34%. The dust impression dataset included impressions deposited on paper and tile, with the latter lifted using either gelatin or Mylar film and an electrostatic lifter. A total of 1,513 RACs were identified from all impressions, generating over 154,000 non-mated RAC comparisons. The RACs in dust impressions were often similar in size or larger than their known mates, leading to a 42% decrease in indistinguishable RAC pairs relative to mated test impressions. As a result, relative RAC-RMFs of at least 0.0008 were observed at a rate between 3.1% and 32%, despite twice the number of RACs available compared to the blood impressions. This contrast suggested that a liquid medium may erode RAC size, while a particulate medium maintains or possibly increases RAC size, thus influencing non-mated RAC similarity. However, no more than one shared indistinguishable RAC pair was observed between unrelated outsoles for either dataset, meaning that an average of four and eight distinguishable RACs were present for blood and dust impressions, respectively.

This research provided estimates of RAC-RMF for a large database of high-quality test impressions as well as two datasets of simulated crime scene impressions. Analysis of these datasets demonstrated that RAC geometries do repeat on non-mated outsoles, and the rate at which this occurs within each dataset was quantified. Since theoretical models have traditionally been the basis for estimating RAC-RMF in footwear and the majority of empirical studies have reported RAC-RMFs at or near zero, the contribution of this research to the forensic footwear community is a calibration of this estimate based on empirical data from a larger sample of outsoles.

Acknowledgements

The database used in this investigation was originally supported by Award No. 2013-DN-BX-K043, awarded by the National Institute of Justice (NIJ), Office of Justice Program, U.S. Department of Justice. The simulated crime scene dataset was supported by the Center for Statistics and Applications in Forensic Evidence (CSAFE), through Cooperative Agreement No. 70NANB20H019 between NIST and Iowa State University, which includes activities carried out at West Virginia University. Images can be found in the CSAFE repository and are cited to Smale, A., Speir, J., West Virginia University 2022 High Quality and Simulated Crime Scene Dataset, Release #1; Release Date March 2022. In addition to the NIJ and CSAFE, thank you to Claire Dolton for registering the 330 simulated crime scene impressions used in this research. Lastly, the opinions, findings, conclusions, and recommendations expressed in this manuscript are those of the authors and do not necessarily reflect those of the Department of Justice, the Center for Statistics and Applications in Forensic Evidence, and/or the National Institute of Standards and Technology.

Contents

1. Introduction

When analyzing footwear impression evidence, one of the goals of an examiner is to determine if a known shoe could be the source of a questioned impression. The two main types of features in footwear impressions that are evaluated during this comparison are class characteristics and characteristics of use. Class characteristics include the make, model, and size of the shoe, which in combination, determine the overall tread pattern. These characteristics are highly consistent between shoes of the same brand and are imparted during the manufacturing process [1]. Randomly acquired characteristics (RACs) can include nicks, gouges, cuts, or foreign material such as rocks or gum deposited on the outsole [1]. These characteristics are typically acquired during everyday wear, as opposed to a direct result of the manufacturing process. However, the spatial location in which these features can develop is dictated by contact area of the shoe, which varies with make and model. During a comparison, a footwear examiner must assess the similarity of class characteristics and RACs between a questioned impression and a known shoe, as well as the rarity of the characteristics that are present. Each determination is based upon the quantity and quality of the observed characteristics, but specific descriptions of what is considered sufficient quantity and quality are not well-defined [2]. As a result, opinions regarding source association are made at the discretion of the examiner based on training and experience.

Due to the purported subjective nature of this interpretation, the opinions formed regarding footwear evidence can be misunderstood, and both under- and overvalued. One way to mitigate this criticism is to complement casework with research that includes quantitative analyses, which has been done successfully within the discipline of DNA as a way to assess the rarity of evidence. Random match probability (RMP) is commonly used in DNA analysis to report the chance that a random person in the population, who is not the donor of the evidence, would have the same DNA profile as observed in the evidence [3]. The aim of this research was to perform a similar evaluation for footwear evidence. However, the term random match "probability" implies extrapolation into a larger population that has not been evaluated. Instead, the term random match "frequency" (RMF) was employed in this investigation to reinforce that the estimates provided are based on direct observations from a database, and no attempt was made to predict the chance of a random match within a non-sampled population. It is also important to note that these estimates are not the same as impression-wide estimates of RMF. When considering a casework scenario, two shoes must share make, model, size, wear patterns, and RACs to be considered a random match. However, in the absence of a large sample of shoes of the same make, model, and size, only the random match frequency of randomly acquired characteristics (RAC-RMF) was evaluated. Make, model, and size were controlled for through an normalization procedure which re-mapped all RACs to a standardized outsole, allowing for RACs in the same relative position on different outsoles to be compared. While this process eliminated some class characteristics (not including areas of tread in contact with the ground), which are necessary to consider during a footwear comparison, it increased the number of RAC comparisons possible in the database.

The majority of previous research has been based on theoretical models[4–6] or empirical studies which have shown that RACs with positional and geometric similarity rarely, if ever, occur on

unrelated outsoles [7–11]. However, the number of possible positions for a RAC to develop is limited by the size of the outsole, and when considering only two-dimensional impressions, the portions of the outsole that have tread in contact with the ground. As a result, the chance of RACs occurring in the same position increases for two unrelated shoes of the same make, model, and size. Likewise, although RACs can vary greatly in shape and size which can lead to more or less complex geometries, RAC features are expected to repeat with sufficient sampling. Thus, it was hypothesized that the findings in [7–11] were the result of limited samples available for analysis rather than the impossibility of similar RACs occurring on unrelated outsoles, therefore prompting further investigation.

While there is no set number of shoes or RACs to define a population that is "large enough" to show repetition of RAC geometries, the chance of observing indistinguishable RAC pairs increases with the size of the database. However, there are several difficulties associated with acquiring a large database for analysis. While it takes time to simply collect a large number of shoes, the time required to process each shoe is much more extensive. Creating high-quality test impressions of the outsoles, marking each shoe for RACs, and comparing the RACs between different outsoles takes a great deal of time and expertise. The West Virginia University (WVU) footwear database [12], which is composed of 1,300 shoes that collectively contain 80,668 randomly acquired characteristics of varying size, shape, and position on the outsole, presented a viable solution to overcome the issue of database size. In addition, past research investigating chance association of RACs [13] demonstrated that indistinguishable RAC pairs with positional similarity occur between non-mated outsoles in this database.

The first phase of this research, which is discussed in detail in Chapter 2, was to estimate RAC-RMF using test impressions of the 1,300 shoes in the WVU footwear database. These test impressions were created in a controlled laboratory setting, resulting in high-quality impressions which allowed for visualization of fine details, including numerous RACs of various sizes. Although these impressions were of much higher quality than the impressions commonly encountered at crime scenes, they provided a useful point of reference for RAC-RMF. Estimation of RAC-RMF required each shoe in the database to be sequentially held out and compared to the remaining 1,299 shoes to determine if any unrelated pairs of shoes share indistinguishable RACs in the same relative position on the outsole. While visual assessment by a human observer is perhaps the most robust measure of indistinguishability known to be available at this time, its most notable drawback is the extensive time required to compare RACs. Alternatively, various automated metrics have been used to assess RAC similarity, including percent area overlap [13], Hausdorff distance [6, 13, 14], Euclidean distance [13, 14], matched filter [14], modified cosine similarity [14], modified phase only correlation [14], and rarity score [15]. With such a large number of RACs available in the WVU database, it was necessary to find a method that effectively compared RACs to determine indistinguishability in a time-efficient manner. Thus, a method that combined visual assessments and predictions from a mathematical model based on a percent area overlap similarity score was implemented, as first described in [13]. The probability of indistinguishability of all non-mated RAC pairs with positional similarity was determined using these methods, and RAC-RMF was reported for each of the 1,300 shoes in the database as the number of unrelated shoes out of 1,299 that shared at least one indistinguishable RAC with the held-out shoe.

After estimating RAC-RMF using test impressions, the next phase of this research was to better understand how RAC-RMF varies with impressions of different quality. Previous research has shown that the majority of RACs do not transfer to crime scene-like impressions [14]. With fewer RACs available in the impressions, the chance of finding indistinguishable RACs shared between unrelated outsoles was expected to decrease. It was also anticipated that the use of different impression media and substrates would affect the size and shape of RACs in the resulting impressions, therefore

increasing or decreasing the degree of similarity between non-mated RACs to an unknown extent. To investigate these possibilities, two different simulated crime scene impression datasets were created to serve as lower-quality complements to the database of high-quality test impressions.

The first dataset, described in Chapter 3, was composed of impressions made in blood on tile and enhanced with leucocrystal violet. This dataset included impressions of varying degrees of totality and of overall lower quality than test impressions, making them more aligned with questioned impressions received in casework. As a result, the influence of a liquid medium on RAC transfer, including number of RACs and RAC size/shape, was investigated. Subsequently, the effect of these variables on RAC-RMF was quantified and compared to the results of the high-quality test impressions. The second dataset, described in Chapter 4, included impressions created in dust. A variety of substrates and collection techniques were employed, including impressions deposited on paper and impressions deposited on tile lifted with either gelatin or Mylar film and an electrostatic lifter. The purpose of this second dataset was two-fold. First, these impressions provided another reference point to compare to the RAC-RMF estimates for the database of high-quality test impressions. Second, the dust impressions were compared to the blood impressions to determine the effect of different media and substrates on RAC transfer and RAC-RMF.

In summary, this research had three main goals. The primary aim was to estimate random match frequency of randomly acquired characteristics for a footwear database of high-quality test impressions. Since RAC-RMF has not previously been investigated in a database of this size, the test impressions analyzed in this study provided a reasonable point of reference. As a continuation, the second goal was to create two simulated crime scene datasets using subsets of shoes from the larger database, and to estimate RAC-RMF for the shoes in these datasets. After analyzing all three datasets, the final goal was to compare the associated RAC-RMF estimates. Overall, this research demonstrated that it is possible to observe indistinguishable RACs on non-mated outsoles, and quantified the rate at which this occurs within datasets of three different impression types. Since theoretical models have traditionally been the basis for estimating RAC-RMF in footwear and the majority of empirical studies have reported RAC-RMFs at or near zero, the contribution of this research to the forensic footwear community is a calibration of this estimate based on empirical data from a larger sample of outsoles.

2. RAC-RMF of High-Quality Test Impressions

This manuscript and the associated supplemental material were published in *Science & Justice*: A. N. Smale, J. A. Speir, Estimate of the random match frequency of acquired characteristics in a forensic footwear database, Science Justice 63 (3) (2023) 427–437. doi: `https://doi.org/10.1016/j.scijus.2023.04.007`. Minor updates were made in July 2023.

Estimate of the Random Match Frequency of Acquired Characteristics in a Forensic Footwear Database

Alyssa N. Smale[a], Jacqueline A. Speir[a,*]

[a]*Forensic & Investigative Science, West Virginia University, 208 Oglebay Hall P.O. Box 6121, Morgantown, WV, 26506, United States*

When analyzing footwear impression evidence, a significant task of the forensic examiner is to determine if a questioned impression could have originated from a known shoe. To form this opinion, examiners typically evaluate the similarity, quantity, and quality of shared class characteristics and characteristics of use. Since these criteria are developed through training and experience, and therefore purported to be subjective in nature, the opinions formed regarding footwear evidence can be misunderstood. One way to mitigate this criticism is to complement casework with research that includes quantitative analyses. The aim of this study was to estimate random match frequency of randomly acquired characteristics (RAC-RMF) in a research database comprised of 1,300 outsoles with more than 80,000 RACs. Based on a combination of visual comparisons (>91,000) and mathematical predictions (>3.8 million), results indicate that 32% of the outsoles in this dataset do not share any indistinguishable RAC pairs with each other, while 19% possess RAC-RMFs of 1 out of 1,299. At the other extreme, the maximum RAC-RMF observed was 49 out of 1,299. These results are based on high-quality test impressions, human assessments, and a single quantitative similarity metric, so they are considered specific to this dataset and method of analysis. Results could differ in other databases and with impressions of lower quality, and therefore should not be extrapolated to casework. Despite this limitation, the results provide a point of reference for how often RACs may repeat in position and geometry on non-mated outsoles, therefore forming the basis for future research.

Keywords: footwear; randomly acquired characteristics; random match frequency; non-mated pairs; percent area overlap

1. Introduction

When analyzing footwear impression evidence, one of the goals of an examiner is to determine if a known shoe could be the source of a questioned impression. In order to form an opinion regarding the possible source, examiners evaluate the similarity of the class characteristics and characteristics of use present in the impressions. The resulting opinion is

*Corresponding author

Email address: Jacqueline.Speir@mail.wvu.edu (Jacqueline A. Speir)

based on the quantity and quality of the observed characteristics, but specific descriptions of what is considered sufficient quantity and quality are not well-defined [1]. As a result, opinions regarding pattern evidence are made at the discretion of the examiner, based on training and experience. Over the course of a career, an examiner can observe countless class characteristics and randomly acquired characteristics (RACs), and will mentally store information about the attributes of these characteristics. Since class characteristics tend to be similar within brands, they result in a smaller number of possible geometric designs to be observed. Alternatively, RACs have a much greater number of possible sizes, shapes, and configurations. It is assumed that with training and experience, examiners continuously update and reinforce their internalized knowledge regarding the likelihood that a given RAC would repeat on unrelated outsoles by chance alone, which informs their judgment concerning the rarity of characteristics as they are observed.

Due to the purported subjective nature of the interpretation of this evidence, the opinions formed regarding footwear evidence can be misunderstood. One way to mitigate this criticism is to complement casework with research that includes quantitative analyses. This has been done successfully within the discipline of DNA as a way to assess the rarity of evidence. The term random match probability (RMP) is most commonly used in forensic DNA analysis to provide an estimate of the chance of randomly selecting a person from a population and observing a predefined DNA profile [2]. In order to report such a statistic, a large number of profiles have been sampled and the resulting frequency information is used to inform predictions regarding the chance of observing similar or different characteristics in additional profiles. This means that while the probability estimate is based partially on empirical data, its widespread use requires some additional assumptions and predictions of what is expected to be valid in non-sampled and unobserved scenarios. With regard to this research, a similar approach was taken to inform a random match frequency (RMF) of randomly acquired characteristics (or RAC-RMF). To compute the random match frequency of RACs, each RAC on every shoe in the chosen database was sequentially compared to RACs with positional similarity on all remaining outsoles. However, no attempt was made to extrapolate these results and predict the random match probability within any other population that has not been observed. As a consequence, the expression "random match frequency" (rather than probability) is employed. It is also important to note that in order for two unrelated outsoles to be considered a random match, they must share make, model, size, wear patterns, and randomly acquired characteristics. Of these five attributes, RACs are the sole focus for this research, without regard for class attributes. As a result, this investigation uses the expression RAC-RMF to describe the random match frequency of the randomly acquired characteristics in an opportunistic dataset not pre-selected to ensure consistency in class characteristics.

To apply the concept of random match frequency of RACs to footwear, it is first necessary to gain an understanding of the empirical rarity of randomly acquired characteristics. In an ideal scenario, each RAC would be considered uniquely shaped and positioned, and the examiner would be able to discriminate these differences. If true, this would make the chance of finding two RACs that are indistinguishable from each other on two unrelated outsoles essentially equal to zero. However, this is not realistic. The number of possible

positions for a RAC to develop on an outsole is limited by the size of the outsole, and when considering only two-dimensional (2D) impressions and RACs based on "removal" of material, the portions of the outsole that have tread in contact with the ground. Thus, the physical size and tread design of an outsole increase the chance of RACs occurring in the same position on two unrelated shoes of the same make, model, and size. Likewise, although RACs can vary greatly in shape and size, leading to more or less complex geometries, RAC features are expected to repeat with sufficient sampling. While it is unlikely for large and complex RACs to repeat, the presence of small and geometrically simple RACs is likely to increase the chance of identifying similar RACs on unrelated outsoles.

Although it is unknown how many shoes must be sampled to observe the repetition of RAC features on unrelated outsoles, the creation of a footwear database is a useful way to obtain a substantial number of RACs for analysis. The West Virginia University (WVU) footwear database [3] is comprised of 1,300 outsoles cataloged by make, model, size, and degree of wear. High-resolution scans (600 PPI) and a corresponding Handiprint exemplar using fingerprint powder were produced for each outsole [3]. Each pair was digitally co-registered by selecting eight common ground control points in both images. Using oblique illumination and 4X magnification to examine the physical outsole, any RACs that appeared both on the outsole and in the test impression were marked on the Handiprint image using the pencil tool in Adobe® Photoshop®. RAC maps were then created by extracting all marked areas, resulting in a binary image that revealed the relative location and shape of each identified RAC. The RAC map for each right shoe was inverted to ensure a common orientation for all shoes. Assuming the center of the image frame as the origin (0,0) of a coordinate system and zero degrees defined by a horizontal axis drawn from (0,0) to the frame's extreme right, a polar coordinate triple (r, r_{norm}, θ) was extracted for each RAC in the RAC map. The radius (r) and angle (θ) represent traditional polar coordinates that localize a RAC's centroid, and the normalized radius (r_{norm}) was obtained by dividing r by the distance from (0,0) to the shoe's perimeter at θ [3]. Next, a Men's size 10 Reebok® walking shoe was divided into 5 mm $\times$ 5 mm bins, generating a total of 987 spatial cells [3]. In the same manner as described above, the polar coordinates and r_{norm} of each spatial cell were extracted. This allowed every RAC in the database to be mapped to one of 987 spatial cells on the Men's size 10 Reebok® walking shoe based on corresponding r_{norm} and θ values [3]. The aim of this normalization procedure was to place RACs that exist in the same relative position on different outsoles (*e.g.*, the upper right toe or the center of the heel) in the same position on a single outsole to allow for comparisons based on positional similarity, regardless of each individual shoe's make, model, or size. Following extraction and localization, each RAC was categorized as either linear, compact or variable in shape [4]. Using this process, a total of 80,668 RACs, extracted from 1,300 outsoles, were available for analysis.

While a large database increases the chance of observing similar RACs on unrelated outsoles, comparing RACs becomes more time-consuming as the size of the database grows. To combat this issue, it is advantageous to implement a comparison method that attains adequate results while maintaining a reasonable level of efficiency. To date, estimating the similarity of two unrelated RACs to determine if they are distinguishable from each other

has been accomplished both mathematically and visually. The use of similarity scores to compare RACs has not been exhaustively investigated, but notable metrics that have been studied include percent area overlap [4], Hausdorff distance [4–6], Euclidean distance [4, 6], matched filter [6], modified cosine similarity [6], modified phase only correlation [6], and rarity score [7]. However, there is great variety in the performance of each metric as these scores analyze RACs from a purely mathematical standpoint, and the full complexity of accidental features cannot always be captured when reduced to a single numerical value.

In terms of visual assessment, Wilson [8] conducted a study to evaluate the similarity of the entire outsole of 39 pairs of shoes of the same make and model with similar levels of wear. All shoes were reported to be distinguishable from each other based on the number, shape, and position of RACs on the outsole. Several other studies have visually compared single features on unrelated outsoles. Cassidy [9] analyzed two different groups of outsoles, one containing 38 shoes of the same make and model and one containing 60 shoes of four different models. The outsoles were compared to search for RACs in the same position on unrelated outsoles, without regard to RAC shape. A 1 in 38 chance of positional similarity was reported for the first group, and a 1 in 60 chance was reported for the second. Adair *et al.* [10] examined 24 outsoles of the same make and model, and found no indistinguishable characteristics. Hamburg and Banks [11] examined RACs on eight different outsoles of the same make and model at different step intervals, and again, reported that all features were distinguishable. Marvin [12] visually assessed RACs on 34 outsoles of the same make and model. Each RAC was compared to other RACs in the same position on different outsoles, requiring 5,551 pairwise comparisons. Again, all RACs with positional similarity were deemed distinguishable from each other.

A chance association study by Richetelli *et al.* [4] implemented a method to efficiently compare a large number of RACs by combining quantitative and visual methods. Using the WVU footwear database, RACs in the same position on unrelated outsoles were compared pairwise using a percent area overlap similarity score, and subsequently ranked based on mathematical similarity. The most similar pairs were visually evaluated by research analysts, resulting in 91,607 comparisons. The conclusions from the visual comparisons were used to inform a mathematical model to assess the similarity of additional RAC pairs based on percent area overlap. This method was implemented to determine the chance that two unrelated outsoles possess similar randomly acquired characteristics in the same position. Along with other summary statistics, the following median probabilities of chance association were reported for each of the three RAC categories: 1 in 444,126 for linear, 1 in 291,111 for compact, and 1 in 880,774 for variable (please see [4] for additional details). This study showed that repetition of indistinguishable RACs in the same position on different outsoles exists within this database, while simultaneously demonstrating the utility of using a method for RAC comparison that combines both quantitative and visual analyses.

In contrast with chance association investigated by Richetelli *et al.* [4], random match frequency/probability addresses the casework-relevant question. At the RAC level, chance association explores how often one might randomly select two unrelated shoes and observe RACs in the same position that are indistinguishable from each other. Conversely, random match frequency involves choosing one RAC and determining the chance of randomly select-

ing a second RAC with positional similarity on an unrelated outsole that is indistinguishable from the first RAC. The hold-one-out concept of RMF provides insight regarding the number of outsoles in a relevant population that could potentially be falsely included as the source of a questioned impression. While exploration of RMF has implications in casework scenarios, the use of high-quality impressions in a research database for the purposes of this study allowed for characterization of the magnitude of RMF with two main caveats. First, these impressions are superior to those commonly encountered at a crime scene. Therefore, this research provides a point of reference for this statistic under controlled laboratory conditions, and the results should not be extrapolated to lower quality impressions without additional study. Second, the results presented here do not represent forensic RAC-RMF in casework scenarios, as make, model, size, and wear patterns were disregarded and only acquired characteristics were evaluated.

2. Materials and Methods

To estimate random match frequency, pairs of RACs in the WVU footwear database were previously compared [4] in order to identify those that deemed "indistinguishable" from each other. For the purposes of this research, indistinguishable was defined as exhibiting minimal pixel-level variation with an allowance for minor variations in length, width, shape, and orientation that could be expected from known mate replicate test impressions [4] (please see supplemental material section S1 and Fig. S1 for examples). Selecting one shoe at a time, each shoe was held out and compared to the remaining 1,299 shoes in the database. Rather than using the exact location of a RAC on the outsole, positional similarity of RACs was determined by a previously implemented binning process [3] based on 5 mm $\times$ 5 mm cells on a normalized outsole. As a result, multiple RACs could exist within the same spatial cell on different outsoles with slight positional differences, but be considered to co-occur in position within this study. In these instances, each RAC was evaluated separately by comparing it to all other RACs in the same cell on all other shoes.

Due to the large number of necessary comparisons, the visual assessments performed by Richetelli *et al.* [4] were re-used for this research. All RAC pairs (i and j) in the same spatial cell on unrelated outsoles were compared using percent area overlap, as shown in Eq. 1 [4]. A maximum of the 25 most similar RAC pairs in each cell per category (linear, compact, and variable) were evaluated by two analysts (note that some cells had fewer than 25 pairs for evaluation). From these visual comparisons, each RAC pair was assigned an indistinguishability value I of either 0.0 or 1.0 for RACs that were deemed distinguishable or indistinguishable from each other, respectively [4]. Upon review, 21 features were deemed to be class, subclass, or manufactured characteristics (*e.g.,* worn-through honeycomb elements) and were eliminated in this analysis, reducing the total number of visual comparisons in this study from 91,607 to 91,600.

$$\%A_{ij} = \frac{[\text{Area of Overlap}] \times 2}{[\text{Area of } i + \text{Area of } j]} \times 100\% \tag{1}$$

For all RAC pairs not visually assessed, the mathematical models implemented by [4] were re-used to estimate the probability of indistinguishability, or the probability that a human examiner would judge the RACs to be indistinguishable. The results of the visual comparisons were modeled as a Bernoulli distribution, $y_{ij} \sim Bernoulli(p_{ij})$, where y_{ij} represents the binary outcome from the comparison and p_{ij} represents the probability of indistinguishability. The conditional probability of indistinguishability based on the percent area overlap of two RACs i and j, or $p(I|\%A_{ij})$, was modeled using a binary logistic regression with a polynomial of degree k as shown in Eq. 2, with the best-fit models occurring when k was equal to four for linear RACs and three for both compact and variable RACs (please see [4] for a discussion of models and fit). If the two RACs were of the same category (linear, compact, or variable), the model describing that category was used for the computation. Conversely, if the RACs under comparison were from two different categories, both corresponding models were used to compute a probability and the higher probability was retained as a worst-case scenario. Probabilities were computed using both the fitted model and the upper 95% credible interval (CrI), with the higher probabilities from the upper credible interval estimate being reported here, again as a worst-case scenario.

$$\begin{aligned} logit(p_{ij}) &= \alpha_0 + \sum_{k=1}^{K} \alpha_k \times (\%A_{ij})^k \\ p_{ij} &= p(I|\%A_{ij}) \end{aligned} \tag{2}$$

Collectively, these two techniques provided a probability of indistinguishability for all possible RAC comparisons. With this information, three main topics were investigated: the occurrence of indistinguishable RAC pairs resulting from quantitative and visual methods, random match frequency of RACs for each shoe in the database, and the occurrence of multiple shared indistinguishable RAC pairs between unrelated outsoles.

First, to characterize overall RAC similarity in the database, the resulting probabilities of indistinguishability from RAC comparisons were binned as follows: {0.0}, (0.0-0.25), [0.25-0.5), [0.5-0.75), [0.75-1.0), {1.0}. The probability threshold of $t = 1.0$ was selected to define *certain* indistinguishability, where two RACs were visually compared and deemed indistinguishable by an analyst. Conversely, lower thresholds of $t \geq 0.75$ and $t \geq 0.5$ were chosen to define *plausible* and *possible* indistinguishable pairs, respectively. These thresholds were used to detect pairs predicted by the mathematical model in addition to those visually compared. Examination of the RACs in indistinguishable pairs for each threshold allowed for trends to be observed with regard to which attributes, such as RAC category and percent area overlap value, are most common in these pairs.

Second, random match frequency of RACs — or RAC-RMF$_{(m|n)}$ — was computed for each shoe in the database, where n represents the number of indistinguishable RACs shared between a pair of unrelated outsoles and m represents an unknown number of distinguishable RAC pairs that were not evaluated. For example, RAC-RMF$_{(m|n \geq 1)}$ was computed by determining the number of non-mated shoes (out of 1,299) with RAC similarity meeting a specific threshold ($t = 1.0$, $t \geq 0.75$, or $t \geq 0.5$) in *at least one* cell (please see supplemental

material section S2 and Fig. S2 for additional clarification and examples). If multiple RACs existed in the same cell on an outsole, the maximum probability of indistinguishability was retained, serving as a worst-case estimate for the cell. Variation in RAC-RMF$_{(m|n \geq 1)}$ as a function of total RAC count on each held-out shoe was also explored.

Finally, the total number of shared indistinguishable pairs between unrelated outsoles was determined in order to report the maximum number of shared pairs observed between unrelated outsoles in this database, or the maximum value of n in the expression RAC-RMF$_{(m|n)}$. This was of interest as a greater number of shared indistinguishable pairs between unrelated outsoles increases the chance of forming an erroneous source association opinion between a questioned impression and a non-mated test impression. Lastly, the RAC spatial distribution across the outsole of these shared pairs was further examined in order to determine if these pairs tend to cluster in proximity and/or location (*e.g.*, toe, heel, etc.).

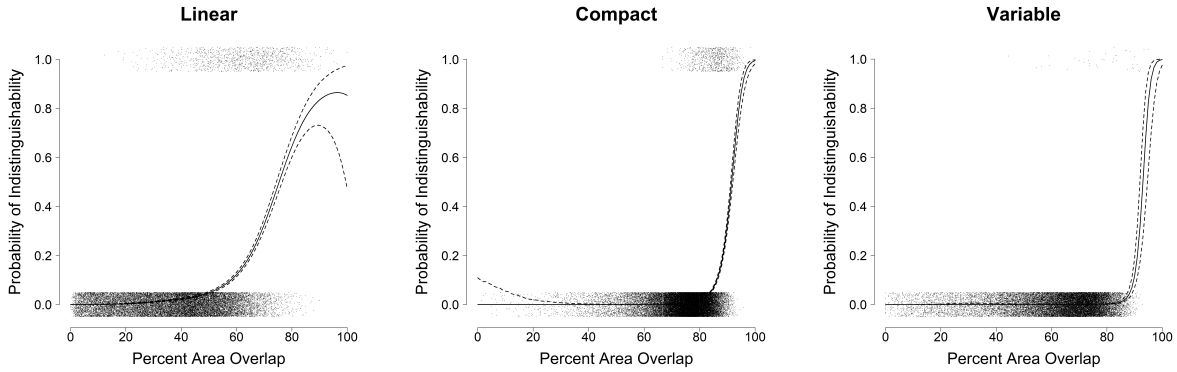
3. Results

3.1. Indistinguishable Pairs

For each pair of shoes, the positions of the RACs on both outsoles were compared to determine if the two shoes shared any RACs in a common location. The maximum number of unrelated shoes containing a RAC with positional similarity to a held-out shoe was 1,290, while the minimum number was 58; in other words, each shoe was compared to at least 58 other shoes containing one or more RACs with positional similarity. Visual inspection or mathematical modeling was used to predict the probability of indistinguishability given a RAC pair's percent area overlap value [4]. The coefficients used for each of the three RAC category fitted models are reported in Table 1. Plots of the probability of indistinguishability based on percent area overlap for each RAC category can be seen in Fig. 1, where the solid line represents the fitted model, the dashed lines represent the 95% credible interval, and the individual data points represent the results of the visual comparisons [4].

Table 1: Fitted model coefficients for Eq. 2 to predict probability of indistinguishability using percent area overlap for each RAC category. Differences from previously published values [4] are attributed to rounding.

	α_0	α_1	α_2	α_3	α_4
Linear	-11.89	0.5254	-0.01411	1.844e-04	-8.228e-07
Compact	-120.6	4.374	-0.05771	2.657e-04	N/A
Variable	-65.42	2.909	-0.04788	2.598e-04	N/A



(a) (b) (c)

Figure 1: Plots of probability of indistinguishability based on percent area overlap for linear (a), compact (b), and variable (c) RAC pairs. The solid line represents the fitted model, the dashed lines represent the 95% CrI, and the data points represent the results of the visual comparisons (jittered vertically ± 0.05). Replotted using raw data; original figures can be found in [4]. Differences in the dashed lines from previously published figures are the result of simulations used to generate the CrI. In other words, the fitted models are consistent, but the credible intervals can vary slightly in locations with higher uncertainty.

Including both visual and predicted comparisons, 3,968,087 comparisons were performed. Of these comparisons, 3,227,855 were previously reported [4]. The difference between this analysis and that previously conducted [4] is the aforementioned removal of class/subclass features deemed to be erroneously included in the dataset, and the addition of 8,383 RACs from 36 of the 1,300 shoes in the existing dataset (please see supplemental material section S3 and Fig. S3 for additional information).

The distribution of probability values for these comparisons is shown in Table 2 for the fitted model and the upper 95% credible interval, both as a count and as a percentage out of the total number of comparisons. Only minor differences were observed between the fitted model and upper credible interval. From this point forward, only the values from the upper 95% CrI will be discussed.

Table 2: Distribution of 3,968,087 probability values resulting from visual comparisons and predictions by both the fitted mathematical model and the upper 95% credible interval [4].

	Fitted Model		**Upper 95% CrI**		**Source**	
Probability	**Count**	**Percentage (%)**	**Count**	**Percentage (%)**	**Visual (%)**	**Predicted (%)**
{0.0}	89,419†	2.25	89,419†	2.25	100	0
(0.0-0.25)	3,875,677	97.7	3,875,532	97.7	0	100
[0.25-0.5)	693	0.0175	800	0.0202	0	100
[0.5-0.75)	100	0.00252	133	0.00335	0	100
[0.75-1.0)	17	0.000428	22	0.000554	0	100
{1.0}	2,181†	0.0550	2,181†	0.0550	100	0

† Values of 89,419 and 2,181 [4] reduced from 89,425 and 2,182, respectively, to account for the removal of the aforementioned elements deemed class or subclass features.

It is clear that the majority of probabilities are close to 0.0, indicating that most RAC pairs were deemed distinguishable by these comparison methods. Using 91,600 of the visual comparisons from [4], there were 89,419 distinguishable pairs assigned a value of 0.0 and 2,181 indistinguishable pairs assigned a value of 1.0 ($t = 1.0$). As a consequence, the probability values of 0.0 and 1.0 consist entirely of the visual results, while the mathematical model is responsible for the values in between these extremes. This is most likely because the visual evaluation of over 91,000 RAC pairs (associated with binary decisions of 0.0 or 1.0) included any RAC with reasonable similarity in geometry (or those that would likely be assigned a probability of indistinguishability of 0.75-1.0 if a continuous scale were permitted). When using a lower threshold of $t \geq 0.75$, 22 additional *plausible* indistinguishable pairs were predicted by the mathematical model for a total of 2,203 pairs. An additional 133 *possible* indistinguishable pairs were predicted at a threshold of $t \geq 0.5$ for a total of 2,336 pairs.

Fig. 2 shows four examples of RAC pairs on unrelated outsoles that were deemed indistinguishable from each other by a human analyst ($t = 1.0$). Note that this judgment does not imply that the features are exactly the same, but that they possess minor variations that could be observed in replicate test impressions — which is the criteria used to define "indistinguishability" in this research study. The majority of indistinguishable RACs were small features with simple geometries (please see supplemental material section S1 and Fig. S1 for additional examples).

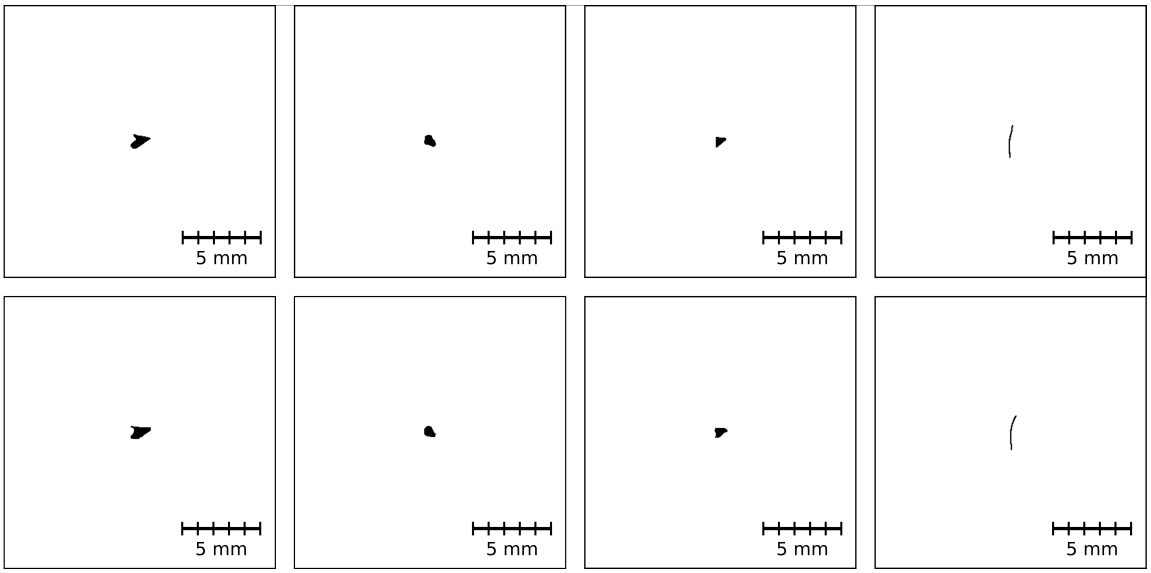
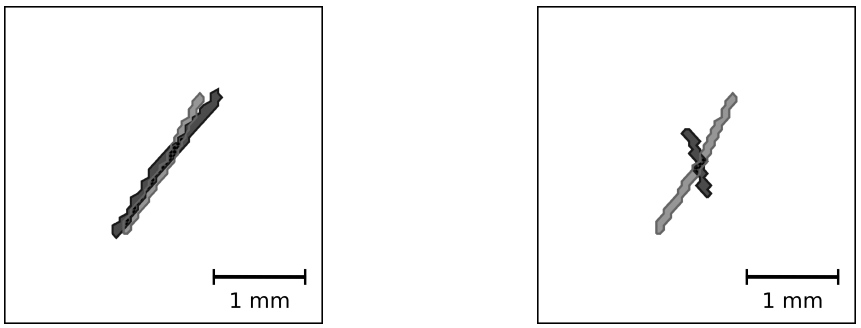


Figure 2: Examples of RAC pairs on unrelated outsoles deemed indistinguishable by a human analyst, where each pair is shown in a column.

Indistinguishable pairs were further explored for each of the three thresholds. For $t = 1.0$, the *certain* indistinguishable pairs ranged in %A overlap from 12.3% to 100%, with an average ± one standard deviation of 71.5% ± 16.0%. For $t \geq 0.75$, the percent area overlap values for the additional 22 *plausible* indistinguishable pairs ranged from 83.1% to 97.9%, with an average ± one standard deviation of 88.6% ± 4.40%. For $t \geq 0.5$, the percent area

overlap values for the additional 133 *possible* indistinguishable pairs ranged from 74.4% to 93.0%, with an average ± one standard deviation of 78.5% ± 4.42%.

The small area overlap value of 12.3% to define an indistinguishable pair requires explanation. Figs. 3(a) and 3(b) depict a single elongated feature (RAC i), measuring 1.7 mm in length and shaded in light gray. The dark gray RAC in Fig. 3(a) (RAC j) measures 2.0 mm in length, and was deemed indistinguishable from RAC i based on $\%A_{ij}$ = 12.3% [4] (indicated by the darkest shaded region). Since the criteria of indistinguishability in [4] was defined as exhibiting minimal pixel-level variation with an allowance for minor variations in length, width, shape, and orientation that could be expected from known mate replicate test impressions, these two RACs qualify. In contrast, the dark gray RAC j in Fig. 3(b) measures 0.8 mm in length, and was deemed distinguishable to RAC i, but likewise has a $\%A_{ij}$ = 12.3% [4]. Thus, a wider range of %A values can be associated with elongated features deemed both indistinguishable (Fig. 3(a)) and distinguishable (Fig. 3(b)). This is further supported by inspection of the plot of indistinguishability versus percent area overlap shown in Fig. 1. In hindsight, rotations of ± 2–3 degrees could have been performed to determine the maximum percent area overlap possible when allowing for slight changes in orientation.



(a) (b)

Figure 3: Two linear RAC pairs with 12.3% area overlap (darkest shaded region). The pair in (a) was deemed indistinguishable by visual comparison, while (b) is clearly a distinguishable pair.

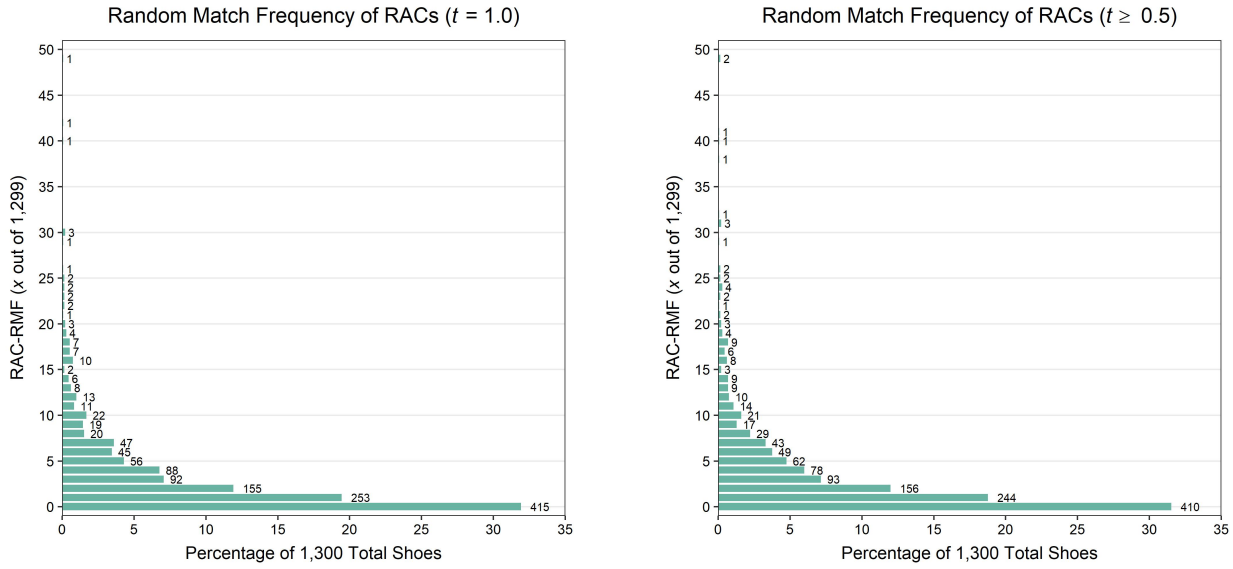
The distribution of RAC pair categories for each threshold is shown in Table 3. The visually-compared indistinguishable pairs used for this study (culled from [4]) were comprised of 1,213 linear, 850 compact, 38 variable, and 80 mixed category RAC pairs. At lower thresholds, the *plausible* and *possible* indistinguishable pairs were often linear or mixed category, and no additional variable pairs were included. Overall, the majority of the indistinguishable pairs (≈ 96%) were comprised of RACs from the same category. Linear RACs were the most common (≈ 57%) in indistinguishable pairs, followed by compact (≈ 40%), and finally variable (≈ 3%). Intuition suggests that isometric RACs would be the most common in indistinguishable pairs. However, since RACs can vary slightly in length and orientation when preparing replicate test impressions, short (1.8 mm ± 1.1 mm) linear RACs were deemed indistinguishable more often than RACs of other categories. This does not suggest that linear features are less discriminating than other types of RACs, but instead, should be regarded as a product of the criteria used in assessing indistinguishability.

Table 3: Number of RAC pairs of each category for $t = 1.0/t \geq 0.75/t \geq 0.5$.

Total Pairs	Linear	Compact	Variable	Mixed
2,181/2,203/2,336	1,213/1,229/1,307	850/852/861	38/38/38	80/84/130

3.2. Random Match Frequency of Randomly Acquired Characteristics

The bar plots shown in Fig. 4 report the RAC-RMF$_{(m|n \geq 1)}$ for $t = 1.0$ (4(a)) and $t \geq 0.5$ (4(b)). As there are only slight differences between plots for $t = 1.0$ and $t \geq 0.75$, plots for $t \geq 0.75$ are not reproduced here. The y-axis reports how often a held-out shoe shared *at least one* indistinguishable RAC pair with another unrelated shoe in the database at the specified threshold. The x-axis reports the percentage of held-out shoes corresponding to each RAC-RMF value. For ease of interpretation, this percentage is also reported as the number of shoes out of 1,300 to the right of each bar. For example, at a threshold of $t = 1.0$, 415 shoes out of 1,300 (32%) did not share an indistinguishable RAC with any other shoe in the database. At the other extreme, one shoe shared at least one indistinguishable RAC pair with 49 other shoes. For $t \geq 0.5$, there are two shoes that shared at least one indistinguishable RAC pair with 49 other shoes. An investigation into the manufacturer of all shoes with at least one indistinguishable pair showed that Nike® is the most common manufacturer (please see supplemental material section S4 and Tables S1 and S2 for additional information). However, this result was anticipated as Nike® is the most common shoe in this database, making up approximately 40% of all shoes.



(a) (b)

Figure 4: RAC-RMF$_{(m|n \geq 1)}$ as a value out of 1,299 for $t = 1.0$ (a) and $t \geq 0.5$ (b). For example, at a threshold of $t = 1.0$, 415 shoes out of 1,300 (32%) did not share an indistinguishable RAC pair with any other shoe in the database (RAC-RMF$_{(m|n \geq 1)}$ of 0 out of 1,299), while 155 shoes (12%) shared at least one RAC pair with two other shoes (RAC-RMF$_{(m|n \geq 1)}$ of 2 out of 1,299).

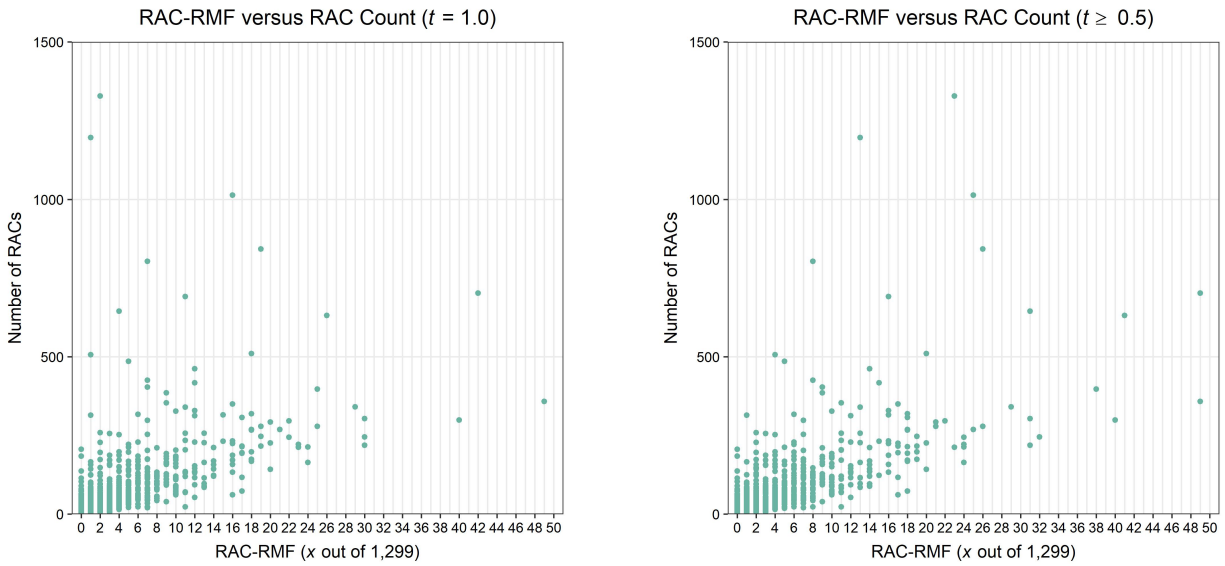
The shoe with the highest RAC-RMF$_{(m|n \geq 1)}$ for $t = 1.0$ was a Men's Route 66® walking sneaker with 358 RACs. This shoe shared one indistinguishable pair with 47 unrelated outsoles and two pairs with 2 unrelated outsoles, totaling 51 shared indistinguishable pairs between this shoe and 49 unrelated outsoles (please see supplemental material section S5 and Fig. S4 for a visual illustration). There were 97 unique RACs in these pairs, 56 of which were compact, 39 of which were linear, and 2 of which were variable in shape. The minimum %A value for these pairs was 47.5%, the maximum was 94.9%, and the average ± one standard deviation was 79.5% ± 11.1%. The pair with the smallest percent area overlap was comprised of two linear RACs, and the low percent area overlap can be explained by a slight difference in angle as previously discussed and illustrated in Fig. 3. The second shoe with a RAC-RMF$_{(m|n \geq 1)}$ of 49 out of 1,299 for $t \geq 0.5$ was a Men's Mozo® shoe with 702 RACs. This shoe shared one indistinguishable pair with 45 unrelated outsoles, two pairs with 3 unrelated outsoles, and three pairs with 1 unrelated outsole, totaling 54 shared indistinguishable pairs between this shoe and 49 unrelated outsoles (please see supplemental material section S5 and Fig. S5 for a visual illustration). There were 103 unique RACs in these 54 pairs, 81 of which were linear, 20 of which were compact, and 2 of which were variable in shape. The minimum %A value for these pairs was 24.6%, the maximum was 96.3%, and the average ± one standard deviation was 69.2% ± 16.0%. The lower percent area values for these pairs can be attributed to the occurrence of linear RACs in the pairs, as well as the lower degree of similarity permitted by $t \geq 0.5$.

The observance of at least one shared indistinguishable pair between these shoes and 49 unrelated shoes is believed to be a function of two main factors. First, the large number of RACs on each of these shoes (a total of 358 and 702) could increase the chance of finding an indistinguishable RAC on another outsole. However, there were several other shoes in the database that had more RACs, but lower RAC-RMF$_{(m|n \geq 1)}$ values, than these shoes. Second, the majority of the RACs in these pairs were small pinpricks or short linear RACs (please see supplemental material section S5 and Figs. S6 and S7 for a semi-random sampling of mated RACs on unrelated outsoles). For the Route 66® shoe, the largest of the compact and variable RACs found to co-occur on another shoe was approximately 0.76 mm wide and the longest linear RAC was 3.4 mm. For the Mozo® shoe, the largest of the compact and variable RACs that co-occurred was approximately 0.64 mm wide, and the longest linear RAC was 3.0 mm. These RACs had simple geometries, which again could increase the chance of finding an indistinguishable RAC on an unrelated pair of shoes.

Although these high RAC-RMF estimates are discouraging in terms of the utility of RACs to accurately differentiate mated and non-mated shoes, they are balanced by two competing factors. First, the RACs in question are small and have simple geometries, which in turn would influence the weight afforded to them by an examiner performing a comparison. Second, with shoes having between 300 and 700 RACs, although the value of m in RAC-RMF$_{(m|n \geq 1)}$ has not been assessed for this dataset, it is reasonable to believe that dissimilarities (or non-mated RACs) on the outsole would be apparent to the examiner, which in turn would influence an examiner's opinion.

The number of RACs on a given outsole varies throughout the database, with a minimum count of one, a maximum of 1,328 (please see supplemental material section S3 and Fig. S3

for additional information), and an average ± one standard deviation of 62 ± 99. For this reason, it was useful to investigate random match frequency as a function of RAC count to determine if the total number of RACs on an outsole had an effect on the RAC-RMF$_{(m|n \geq 1)}$. The scatter plots shown in Fig. 5 display this data for $t = 1.0$ (5(a)) and $t \geq 0.5$ (5(b)), where the x-axis reports the RAC-RMF value out of 1,299 and the y-axis reports the total number of RACs on the held-out shoe. The number of individual data points for each x-value corresponds to the number of shoes with a given RAC-RMF, shown to the right of each bar in Figs. 4(a) and 4(b). There may be a slight trend suggesting that shoes with fewer RACs tend to have a lower random match frequency than those with a greater number of RACs. This is likely explained by the fact that as the number of RACs increases, the chance of finding a similar RAC in the same position on another outsole also increases. However, there are several shoes with 300 to 400 RACs that have very different RAC-RMFs, ranging from 1 out of 1,299 to 49 out of 1,299. This indicates that total number of RACs on a shoe is at most a weak predictor of the maximum possible RAC-RMF for a given shoe. Perhaps the region of most interest is the cluster of points in the bottom left corner. In a crime scene impression, it is anticipated that a relatively small number of RACs will be transferred from the outsole to the impression as a function of variation in medium, substrate, and wearer-activities. Focusing on the shoes with less than 100 RACs, the maximum RAC-RMF value is 17 out of 1,299 for $t = 1.0$ and 18 out of 1,299 for $t \geq 0.5$, with over 90% of values being 5 out of 1,299 or less.



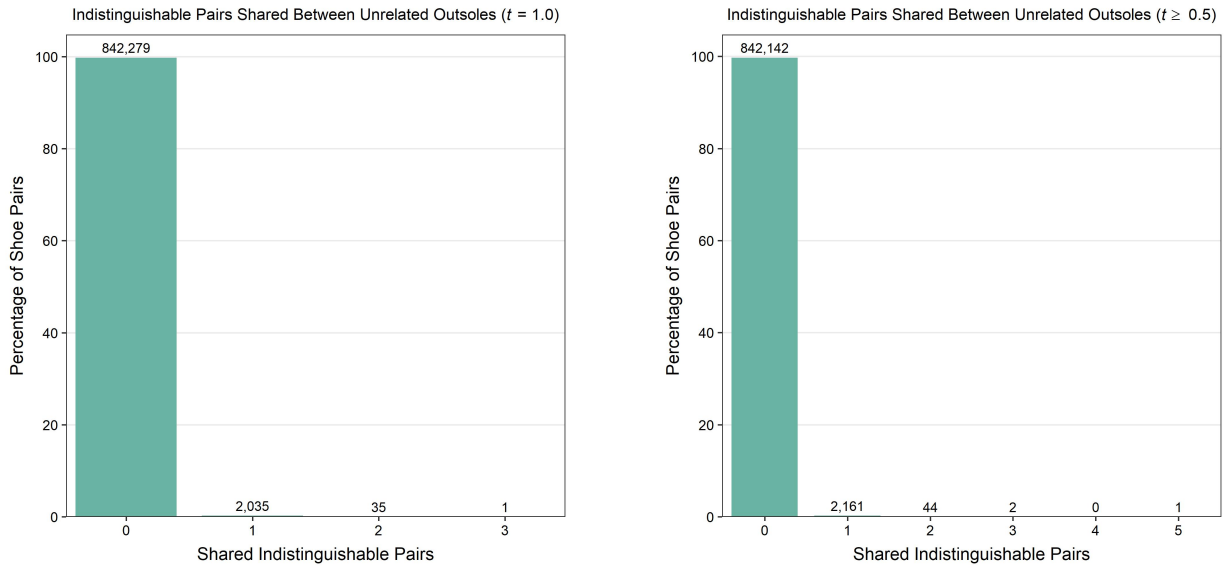
(a) (b)

Figure 5: RAC-RMF$_{(m|n \geq 1)}$ as a function of total RAC count per shoe for $t = 1.0$ (a) and $t \geq 0.5$ (b).

3.3. Multiple Shared Indistinguishable Pairs

Further analysis was performed to explore the occurrence of multiple indistinguishable RACs on a pair of unrelated outsoles in order to find the maximum value of n in RAC-

$\text{RMF}_{(m|n)}$ for this database at the three thresholds of $t = 1.0$, $t \geq 0.75$, and $t \geq 0.5$. The 1,300 shoes can be paired in 844,350 different combinations. Each pair of shoes with a single RAC pair deemed indistinguishable based on the chosen threshold was sequentially selected to determine the number of additional RAC pairs that may be shared between these outsoles. The bar plots in Fig. 6 show the distribution of shoe pairs with each number of shared RAC pairs for thresholds $t = 1.0$ (6(a)) and $t \geq 0.5$ (6(b)). At a threshold of $t = 1.0$, 842,279 pairs of shoes (99.8%) do not share any RAC pairs deemed indistinguishable by visual comparison. Conversely, 2,071 pairs of shoes (0.2%) share one or more indistinguishable pairs, with three being the maximum number of shared indistinguishable pairs ($n = 3$), which was observed on one pair of outsoles (1.2e-04%). At the lower threshold of $t \geq 0.5$, 842,142 pairs of shoes (99.7%) did not share any *possibly* indistinguishable pairs. The remaining 2,208 pairs of shoes (0.3%) share one or more *possibly* indistinguishable pairs, with five being the maximum number of shared RAC pairs ($n = 5$) observed on one pair of outsoles (1.2e-04%). The occurrence of multiple indistinguishable pairs between unrelated outsoles increases the chance of forming an incorrect source association opinion, and thus these pairs were further explored.



(a) (b)

Figure 6: Maximum number of indistinguishable RAC pairs shared between 844,350 unrelated pairs of shoes for $t = 1.0$ (a) and $t \geq 0.5$ (b). The maximum number of shared RAC pairs was three for $t = 1.0$ and five for $t \geq 0.5$. Technically, if a shoe has five shared RAC pairs, it also has four, three, two, and one shared pairs, but this figure reports the maximum (rather than the cumulative) number of shared RAC pairs.

For $t = 1.0$ and $t \geq 0.75$, there were 23 occurrences (out of 2,071 for $t = 1.0$ and 2,091 for $t \geq 0.75$ or $\approx 1.1\%$) of a left and right shoe from the same pair sharing one or more indistinguishable RAC pairs. There were 24 occurrences (out of 2,208 or $\approx 1.1\%$) of left-right pairs for $t \geq 0.5$. Although non-mates, as members of the same pair these shoes are not entirely unrelated. The distribution of these values is shown in Table 4 for $t = 1.0$, $t \geq$

0.75, and $t \geq 0.5$. The percentage of these left-right pairs out of the total number of shoe pairs for each number of shared indistinguishable pairs is also reported. For all thresholds, 20 left-right pairs shared one indistinguishable RAC pair and one left-right pair shared three indistinguishable RAC pairs. For $t = 1.0$ and $t \geq 0.75$, two left-right pairs shared two RAC pairs, while three left-right pairs shared two RAC pairs at $t \geq 0.5$. The shoes with five shared RAC pairs at $t \geq 0.5$ were not a left-right pair. The RAC-RMF values for the shoes in these left-right pairs ranged from 1 out of 1,299 to 49 out of 1,299, with an average of 13 out of 1,299 for $t = 1.0$ and $t \geq 0.75$, and 15 out of 1,299 for $t \geq 0.5$. Overall, the majority of shoes with multiple shared indistinguishable pairs are truly unrelated and are not left-right pairs. In addition to mated left-right pairs, the overall contribution of left and right shoes to indistinguishable pairs was evaluated. When considering all shoes that shared at least one RAC pair with an unrelated shoe, 50.3% were left shoes and 49.7% were right shoes, indicating that a left or right shoe does not increase the chance of finding an indistinguishable pair.

Table 4: Number and percentage of indistinguishable pairs shared between left and right shoes from the same pair for $t = 1.0/t \geq 0.75/t \geq 0.5$.

Shared RAC Pairs	Left-Right Pairs	Total Shoe Pairs	Percentage (%)
1	20/20/20	2,035/2,054/2,161	0.98/0.97/0.93
2	2/2/3	35/35/44	5.7/5.7/6.8
3	1/1/1	1/2/2	100/50/50
4	–/–/0	–/–/0	–/–/0
5	–/–/0	–/–/1	–/–/0

The spatial distribution of single and multiple shared indistinguishable pairs across the outsole was further explored. Table 5 reports the number of cells (out of 987) which contain one or more (to the maximum observed of five) pairs of indistinguishable RACs. Using the second row of the table as an example, for $t = 1.0$ there were 35 pairs of shoes in the database that shared two indistinguishable RAC pairs between them, resulting in a total of 70 RAC pairs in this group. These 70 RAC pairs were distributed in 63 out of 987 cells on the outsole, indicating that at least some of these 63 cells contained more than one pair.

Table 5: Distribution of shared RAC pairs in 987 cells across the outsole for $t = 1.0/t \geq 0.75/t \geq 0.5$.

Shared RAC Pairs	Total Shoe Pairs	Total RAC Pairs	Unique Cells
1	2,035/2,054/2,161	2,035/2,054/2,161	710/710/717
2	35/35/44	70/70/88	63/63/79
3	1/2/2	3/6/6	3/6/6
4	–/–/0	–/–/0	–/–/0
5	–/–/1	–/–/5	–/–/5

The heatmaps in Fig. 7 show the spatial distribution of these RAC pairs across the outsole. Since the distributions have minute differences for the thresholds of $t = 1.0$ and $t \geq 0.75$, the heatmaps for $t \geq 0.75$ are not shown. The heatmaps are similar at all three thresholds for one and two shared pairs, and since the threshold of $t \geq 0.5$ represents *possible* (rather than probable or definite) indistinguishability, only the heatmaps for $t = 1.0$ are shown for these groups (7(a) and 7(b)). Continuing the example using the second row of Table 5, the second image (7(b)) shows the 63 cells in which the 70 RAC pairs are located, and shades of blue indicate the frequency of these pairs in each cell. For the three and four shared pairs groups, heatmaps for $t \geq 0.5$ are shown (7(c) and 7(d)). The circled cells in the plot of the three shared pairs (7(c)) indicate the locations of the three pairs shared between one pair of outsoles at all three thresholds. Conversely, the three cells in the toe are the locations of the three RAC pairs shared between a second pair of outsoles detected at $t \geq 0.75$ and $t \geq 0.5$. The five shared pairs plot for $t \geq 0.5$ (7(d)) shows that the five RAC pairs are each found in a different cell on the outsole.

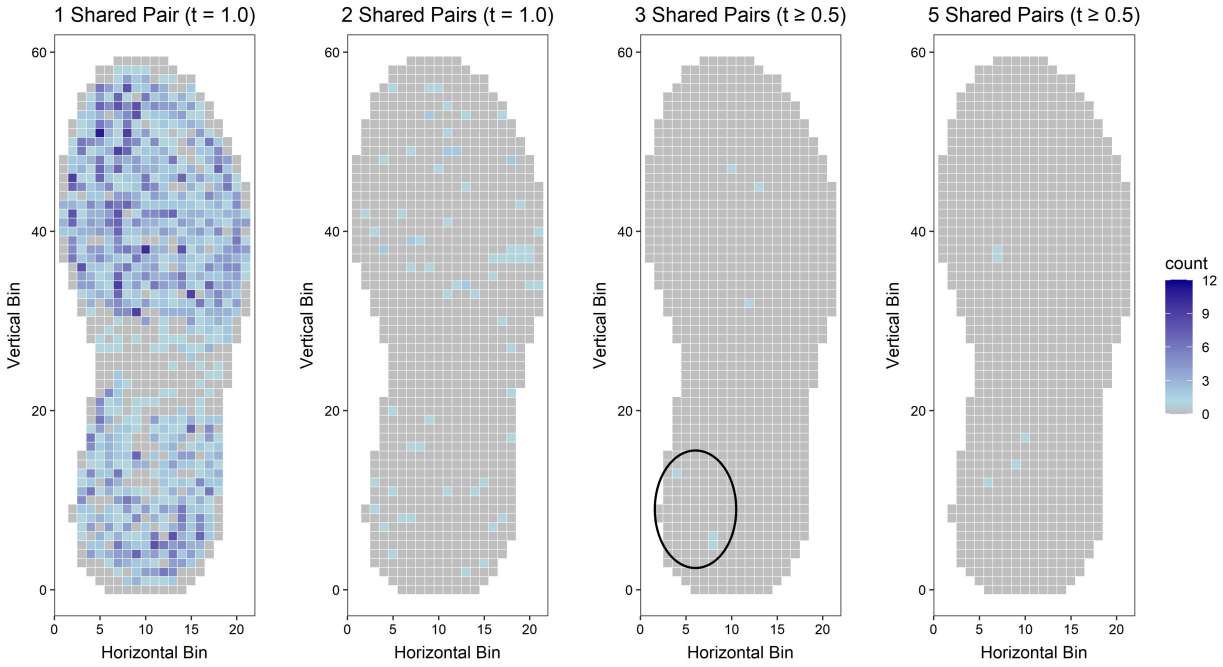


Figure 7: Heatmaps showing the distribution of shared indistinguishable pairs for unrelated shoe pairs that share 1 (a), 2 (b), 3 (c), or 5 (d) RAC features at $t = 1.0$ or $t \geq 0.5$. The circled cells in (c) indicate locations of RAC pairs detected at all three thresholds.

The reason for investigating the distribution of pairs across the outsole is particularly relevant when encountering partial impressions. For example, if two, three, or even four indistinguishable pairs are found to cluster in a specific location on the outsole (such as the ball of the toe or the heel), then the propensity to associate unrelated partial impressions with a known n (where $n = 1, 2, 3$, etc.), but an unknown and reduced m (reduced by

the impression's lack of totality), would be greater than if clustering was not apparent. Interestingly, inspection of Figs. 7(c) and 7(d) suggests that highlighted cells do appear to cluster (more so than random or dispersed). However, spatial autocorrelation was not computed due to the small sample size, but the possibility should be investigated in future work.

4. Discussion

The aim of this study was to investigate the magnitude of random match frequency of RACs within an opportunistic footwear database. For this research, only high-quality test impressions and randomly acquired characteristics were examined. In order to investigate RACs on outsoles that varied in make, model, size, and wear, accidentals were mapped to a standardized outsole through a normalization procedure and localized to 5 mm $\times$ 5 mm spatial cells to determine which RACs had positional similarity on unrelated outsoles (see [3] for complete details). Due to this normalization process, the results reported here suffer from two major issues. First, the make, model, size, and wear patterns of each shoe were disregarded. This is problematic since the practical forensic comparison of accidental features is preceded by class association. This further means that the results do not account for the fact that all shoes do not share equivalent tread in contact with the ground/terrain. Since this study is based on two-dimensional exemplars, each shoe does not have the equivalent potential to develop a RAC. For example, consider comparing a shoe with a raised instep to one with a flat outsole; although a "comparison" is being conducted, RACs in the instep of the second shoe cannot co-occur in position with any RACs on the instep of the first shoe due to the differing outsole designs. Within this study, this is numerically equivalent to a RAC-RMF of 0 out of 1 for cells within the instep — but by virtue of a numerical impossibility rather than a true and observed lack of correspondence. As a result, the mixed makes and models compared in this dataset represent a significant study limitation, and shoes of the same make, model and size may very well have larger RAC-RMFs than that reported here. Unfortunately, acquiring a sufficiently large dataset of shoes of the exact same size and design is equally problematic. In addition to the mixed make and model problem, the second error introduced by the normalization process is the rigid division of RACs into cells or bins wherein there may be instances of RACs in close proximity but on either side of a cell boundary that were not compared, while RACs that were farther apart but within the same cell were compared. Thus, a change in spatial sampling, and/or the sampling origin, could shift the RACs into different bins and generate alternative RAC pairs for evaluation.

As a consequence of these limitations, the results presented here do not represent RAC-RMF in casework scenarios, characterized by lower-quality questioned impressions and RACs on shoes with consistent make, model, size, and wear. Despite these known limitations, the research is still considered valuable as it provides a point of reference estimate of RAC-RMF that is greater than zero, in contrast to past research limited by sample size. With this in mind, comparison of the RACs in this database allowed for an investigation into three main research areas with forensic implications. For convenience, each topic and the relevant results are summarized separately.

4.1. Indistinguishable Pairs

The first goal was to determine how often indistinguishable pairs exist between known non-mated outsoles in this database. Nearly four million RAC comparisons were performed by means of visual assessment or prediction using a mathematical model [4]. Almost 90% of the RAC pairs with positional similarity had a probability of indistinguishability less than 0.05 based on shape characteristics, indicating that most RACs were distinguishable from each other. This probability corresponds to percent area overlap values of approximately 50% or less for linear, 84% or less for compact, and 88% or less for variable RACs. The overall lower %A values observed for linear RACs are the result of their elongated shape, which causes percent area overlap to decrease drastically with a slight change in orientation, despite two RACs appearing similar to a human observer. Conversely, for compact and variable RACs, even pairs with high percent area overlap were often deemed distinguishable.

Despite the fact that the majority of RACs were distinguishable from each other, RAC features were found to repeat within this database. A total of 2,181 RAC pairs were deemed indistinguishable by visual comparison ($t = 1.0$), an additional 22 were deemed *plausibly* indistinguishable by mathematical modeling ($t \geq 0.75$), and an additional 133 were deemed *possibly* indistinguishable ($t \geq 0.5$). These indistinguishable pairs comprise less than 0.1% of all RAC pairs. Of these indistinguishable RAC pairs, the majority were comprised of two RACs from the same class, with linear features being the most common. However, these linear features have an average length ± one standard deviation of 1.8 mm ± 1.1 mm, and therefore would be afforded weight commensurate with their size if observed in a forensic footwear comparison.

4.2. Random Match Frequency of Randomly Acquired Characteristics

The second goal was to determine the maximum random match frequency of RACs observed in this database. RAC-RMF$_{(m|n)}$ was computed for each shoe to determine the number of unrelated shoes that possessed similar features in the same locations as the chosen shoe. The expression RAC-RMF$_{(m|n)}$ was used to reflect the number of shared indistinguishable RAC pairs n, and to highlight that there is an unknown number m of distinguishable pairs that have not been evaluated. This is important in the case of two impressions created by two different shoes that appear to share a feature, but possess some number of other dissimilar features, wherein the total weight of all similarities and differences would be considered before forming an opinion regarding possible source associations. In other words, the unknown number of m dissimilar features matters when forming an opinion, but is not reported in this study.

Using a threshold of $t = 1.0$, it was concluded that 32% of the 1,300 outsoles in this database did not share any indistinguishable RAC pairs with unrelated outsoles ($n = 0$). An additional 19% of outsoles had a RAC-RMF$_{(m|n \geq 1)}$ of 1 out of 1,299. The maximum random match frequency observed in this database was 49 out of 1,299, which occurred for one outsole. This shoe had 358 RACs and shared 51 indistinguishable pairs with 49 unrelated outsoles, the majority of which were comprised of small compact or linear features. When using a lower threshold of $t \geq 0.5$, a second shoe had a random match frequency of 49 out of 1,299. This shoe had 702 RACs and shared 54 indistinguishable pairs with 49 unrelated

outsoles, which were comprised mostly of small linear or compact features. Although RAC-RMF$_{(m|n \geq 1)}$ values near 1 in 1,000 exist for this database, the majority of RAC pairs were distinguishable from each other, and the indistinguishable features were often small RACs that did not possess remarkable attributes. Moreover, these results are based on high-quality impressions capable of reproducing hundreds of identified RACs, and without regard for the number of m different features that may simultaneously exist.

4.3. Multiple Shared Indistinguishable Pairs

The last goal was to determine how often unrelated outsoles share multiple indistinguishable RAC pairs. At least one shared RAC pair was observed on approximately 0.2% of the shoe pairs in this database. At thresholds of $t = 1.0$ and $t \geq 0.75$, a maximum of three shared RAC pairs ($n = 3$) were observed on the 844,350 pairs of non-mated outsoles. At a threshold of $t \geq 0.5$, a maximum of shared five indistinguishable RAC pairs ($n = 5$) were observed. While five shared pairs may be of concern, this occurred at a low threshold of indistinguishability and only on a single pair of unrelated outsoles. Some of the shoes containing one or more shared RAC pairs were the left and right shoe from the same pair, but no strong trend was observed to suggest that a left-right pair increased the chance of finding indistinguishable RACs. The spatial distribution of the multiple pairs showed that these RAC pairs may cluster, and often occur in areas of highest RAC density.

At the higher thresholds, no more than three indistinguishable pairs were ever observed on two unrelated high-quality test impressions, for a maximum value of $n = 3$ in the expression RAC-RMF$_{(m|n)}$. Of course, there is no scientific basis to demand that an opinion of "identification" (or strong support for source association) be limited to outsoles with $n \geq 3$ and $m \leq x$. Naturally, the size, complexity, geometry, and quality of a RAC pair that is in agreement (or disagreement) certainly matters too. However, this study provides a baseline for future investigation. For example, does this persist in datasets of shoes of the exact same make, model, and size? Moreover, is it observed when comparing high-quality test impressions to lower-quality real or simulated crime scene impressions?

4.4. Summary

In conclusion, this research led to the estimation of random match frequency of randomly acquired characteristics for the WVU footwear database. Considering that random match frequency has not been widely investigated in footwear evidence, the use of high-quality test impressions provides a point of reference to begin to understand the magnitude of RAC-RMF. As previously discussed, most past investigations on random association of RACs have failed to detect its occurrence [8, 10–12]. However, this outcome is believed to be a direct result of the limited availability of samples for analysis. In contrast, the dataset and method of analyses conducted here alleviated the issue of inefficiency in RAC comparison across a large number of outsoles, revealing a baseline rate of possible random match frequencies for randomly acquired characteristics. Nevertheless, the reader is cautioned that the results presented here are directly relevant to this research database, and could differ in other databases, with impressions of lower quality, and if using alternative human observers

or similarity metrics. Thus, these summary statistics cannot be directly translated into casework estimates of random association of acquired characteristics.

Despite this caveat, the results have practical relevancy for examiner training and self-calibration when forming subjective opinions, particularly with regard to two major conclusions. First, the possibility of RAC-RMF$_{(m|n \geq 1)}$ was reported at a rate of 0.2% for all shoe pairs evaluated in this dataset. Although this metric should be independently evaluated by other researchers using different methods and alternative datasets, this rate can be considered a first-order estimate of the frequency that non-mated shoe pairs may exhibit at least one RAC with positional and geometric similarity. Moving forward, the authors anticipate three competing factors that are likely to influence RAC-RMF estimates in casework. First, crime scene impressions typically reproduce fewer RACs than a high-quality test impression of the same outsole. In addition, these RACs are expected to be limited to those of larger size (*i.e.*, those reliably detected must be large or prominent enough to exist above the limit of noise introduced by substrate, media, and/or physical activities). Together, these two factors are likely to decrease RAC-RMF estimates in casework scenarios. Lastly, when RAC-RMF is computed using shoes that have a consistent make, model, and size, a pre-existing class association will ensure consistent tread contact, therefore potentially increasing the number of RACs that co-occur in position. In contrast with the number and size of the features reproduced, this third factor is likely to increase RAC-RMF estimates. As a result, the impact of these competing factors on the final statistic cannot be predicted, and caution must be exercised when applying the current rate to casework. However, in the absence of any other empirical studies reporting RAC-RMFs as large as those reported here, examiners can self-calibrate at a possible rate of approximately 0.2% until additional studies allow for this metric to be independently repeated, verified, and/or amended. Of course, this does not mean an erroneous source association will occur at a rate of 0.2%. In other words, two shoes are rarely associated based only on the similarity of a single RAC, but rather on all possible similarities, explainable dissimilarities, and the absence of observable differences (or the value of m in RAC-RMF$_{(m|n)}$ as presented in this study). That said, it is interesting to note (although correlation does not mean causation) that footwear black box studies in recent years involving examiners report false positive rates for source association of 0.2% [13] and 0.48% [14].

The second major conclusion is that no more than three shared RAC pairs were observed in this dataset at indistinguishability levels of $t = 1.0$ and $t \geq 0.75$. Consider the work by Miller and McLean [15] aimed at counting the maximum number of consecutive matching striae on bullets discharged from different firearms. This study concluded that groups of more than six consecutive matching striae in two-dimensional impressions or four in three-dimensional impressions were never observed in non-mated impressions. In a similar manner, the present research can be practically applied; in the absence of any known dissimilarities, four RACs of limited size and geometric complexity have not yet been observed to co-occur by chance on two unrelated outsoles. This dataset allows for self-calibration that the possibility of three or less indistinguishable RAC pairs exists by chance on non-mated outsoles, and therefore when presented with $n \leq 3$, the size, complexity, geometry, and quality of RAC pairs in agreement must be thoroughly considered when forming an opinion of association.

4.5. Future Considerations

Future studies aim to characterize RAC-RMF when presented with simulated crime scene impressions more closely aligned with those expected in casework scenarios. To accomplish this, a subset of shoes from the WVU footwear database will be used to create impressions in blood and dust. Blood impressions will be created on a variety of tile types and chemically enhanced using leucocrystal violet. Dust impressions will either be created on paper, or on tile and lifted with an electrostatic or a gelatin lifter; as necessary, dust impressions will be digitally enhanced using Adobe® Photoshop®. Each simulated crime scene impression will be registered to its corresponding high-quality test impression. Whenever possible, a semi-automated procedure based on ground control points will be used for image registration. However, impression partiality is likely to dictate the need for manual registration.

It is expected that the media and substrate of the crime scene impressions will influence the number of RACs that reproduce as well as the size of those RACs that can be reliably identified. The previously-marked test impressions will be used to assist in marking through a modified sequential unmasking process which aims to balance objectivity and efficiency. It is anticipated that this will result in some biased RAC geometries. Thus, each RAC identified in the simulated crime scene impressions will be marked in Adobe® Photoshop® and tagged with a label denoting when during the unmasking process it was detected. Finally, RAC maps will be created and processed in the same manner as the corresponding test impressions, including localization of RACs to the same standard outsole shape and size used to normalize all RACs extracted from the high-quality test impressions.

For each simulated crime scene impression, three attributes will be evaluated. First, the percentage of RACs that reproduce as a function of the total number of RACs present on the corresponding high-quality test impression will be reported. It is anticipated that this will be much less than 25%, but will vary as a function of the quality of the simulated crime scene impressions. Second, the RACs in the simulated crime scene impressions will be compared to RACs with positional similarity in test impressions made by unrelated outsoles. Similar analyses as presented herein will be performed to determine the number of indistinguishable pairs found and the RAC-RMF of each shoe in the simulated crime scene dataset. Third, the similarity of each RAC from the simulated crime scene impressions and its mated RAC from the corresponding test impression will be evaluated to estimate the degree of change in size, shape, and geometry that may result. The cumulative outcome will be to provide summary statistics to compare simulated crime scene impressions to higher-quality test impressions, thereby elucidating the collective impact of substrate, media, and quality on the number, size, and shape of RACs that are reproduced in simulated casework scenarios, and therefore a shoe's respective RAC-RMF.

Acknowledgments

This dataset used in this investigation was originally supported by Award No. 2013-DN-BX-K043, awarded by the National Institute of Justice, Office of Justice Program, U.S. Department of Justice. However, the opinions, findings, conclusions, and recommendations expressed in this manuscript are those of the authors and do not necessarily reflect those

of the Department of Justice. In addition to the NIJ, the authors would like to thank Dr. Hariharan K. Iyer for his thoughtful feedback on the manuscript.

References

[1] SWGTREAD, Range of Conclusions Standard for Footwear and Tire Impression Examinations, https://www.nist.gov/system/files/documents/2016/10/26/swgtread_10_range_of_conclusions_standard_for_footwear_and_tire_impression_examinations_201303.pdf, 2013 (accessed December 2022).

[2] National Research Council: Committee on DNA Forensic Science, The Evaluation of Forensic DNA Evidence, National Academy Press, Washington, D.C., 1996.

[3] J. A. Speir, N. Richetelli, M. Fagert, M. Hite, W. J. Bodziak, Quantifying randomly acquired characteristics on outsoles in terms of shape and position, Forensic Science International 266 (2016) 399–411. doi:https://doi.org/10.1016/j.forsciint.2016.06.012.

[4] N. Richetelli, W. J. Bodziak, J. A. Speir, Empirically observed and predicted estimates of chance association: Estimating the chance association of randomly acquired characteristics in footwear comparisons, Forensic Science International 302 (2019) 1–14. doi:https://doi.org/10.1016/j.forsciint.2019.05.049.

[5] Y. Yekutieli, Y. Shor, S. Wiesner, T. Tsach, Expert assisting computerized system for evaluating the degree of certainty in 2D shoeprints, Technical Report. National Institute of Justice, U.S. Department of Justice, Office of Justice Programs (2012).

[6] N. Richetelli, M. Nobel, W. J. Bodziak, J. A. Speir, Quantitative assessment of similarity between randomly acquired characteristics on high quality exemplars and crime scene impressions via analysis of feature size and shape, Forensic Science International 270 (2017) 211–222. doi:https://doi.org/10.1016/j.forsciint.2016.10.008.

[7] S. Wiesner, Y. Shor, T. Tsach, N. Kaplan-Damary, Y. Yekutieli, Dataset of digitized RACs and their rarity score analysis for strengthening shoeprint evidence, Journal of Forensic Sciences 65 (3) (2020) 762–774. doi:https://doi.org/10.1111/1556-4029.14239.

[8] H. D. Wilson, Comparison of the individual characteristics in the outsoles of thirty-nine pairs of Adidas Supernova Classic shoes, Journal of Forensic Identification 62 (3) (2012) 194–203.

[9] M. J. Cassidy, Footwear Identification, Public Relations Branch of the Royal Canadian Mounted Police, Ottawa, Ontario, 1980.

[10] T. W. Adair, J. LeMay, A. McDonald, R. Shaw, R. Tewes, The Mount Bierstadt study: An experiment in unique damage formation in footwear, Journal of Forensic Identification 57 (2) (2007) 199–205.

[11] C. Hamburg, R. Banks, Evaluation of the random nature of acquired marks on footwear outsoles, Impression and Pattern Evidence Symposium; Clearwater, FL (2010).

[12] M. Marvin, A look at close non-match footwear examinations, International Association for Identification (IAI) Centennial Conference; Cincinnati, OH (2015).

[13] R. A. Hicklin, B. C. McVicker, C. Parks, J. LeMay, N. Richetelli, M. Smith, J. Buscaglia, R. S. Perlman, E. M. Peters, B. A. Eckenrode, Accuracy, reproducibility, and repeatability of forensic footwear examiner decisions, Forensic Science International 339 (2022) 1–16. doi:https://doi.org/10.1016/j.forsciint.2022.111418.

[14] N. Richetelli, L. Hammer, J. A. Speir, Forensic footwear reliability: Part III—Positive predictive value, error rates, and inter-rater reliability, Journal of Forensic Sciences 65 (6) (2020) 1883–1893. doi:https://doi.org/10.1111/1556-4029.14552.

[15] J. Miller, M. McLean, Criteria for identification of toolmarks, AFTE Journal 30 (1) (1998) 15–61.

Estimate of the Random Match Frequency of Acquired Characteristics in a Forensic Footwear Database

Supplemental Material

S1. Indistinguishable Definition

For the purposes of this research, "indistinguishable" was defined as exhibiting minimal pixel-level variation with an allowance for minor variations in length, width, and orientation that could be expected from known mate replicate test impressions. Fig. S1 shows a selection of RAC pairs that were deemed indistinguishable by a human analyst. Additional examples can be found in Richetelli *et al.* (2019).

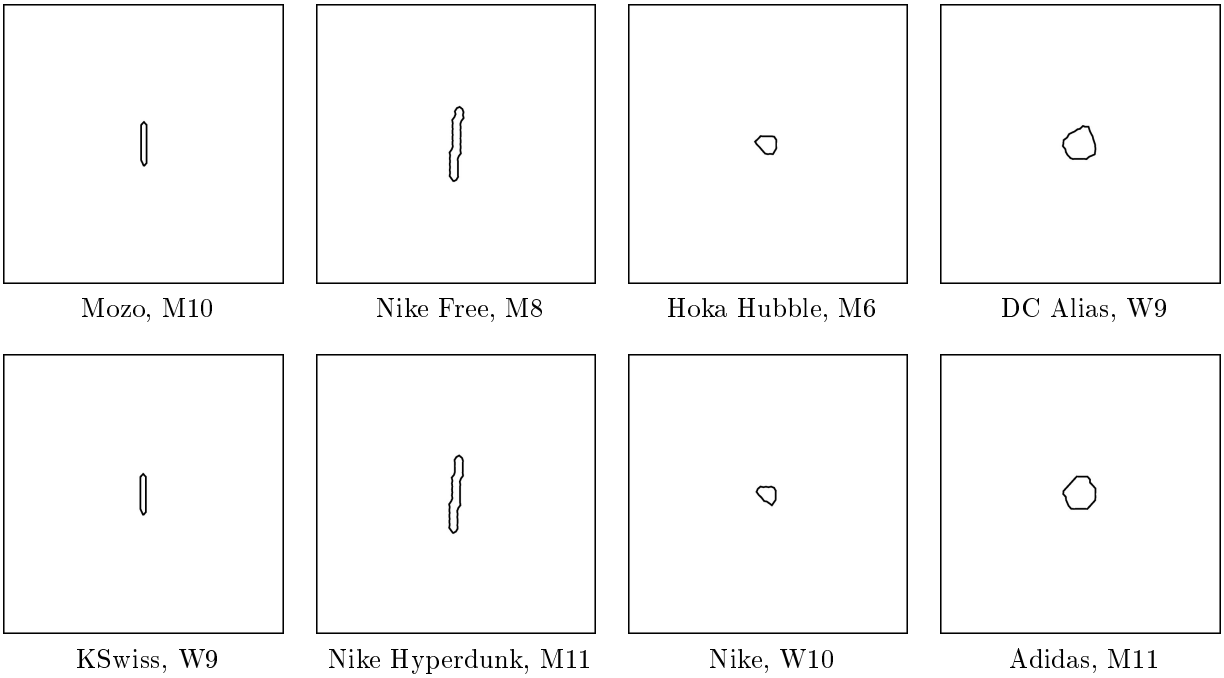


Figure S1: Examples of RAC pairs that were visually deemed indistinguishable by a human analyst using the criteria outlined in Richetelli *et al.* (2019). Each pair is shown in a column. The make, model (when available), and size (Men's (M)/Women's (W)) of the shoe containing the RAC are also reported. Registered trademark symbols are omitted for brevity in expression.

S2. RAC-RMF Notation

The notation RAC-RMF$_{(m|n)}$ was used to report the random match frequency of RACs. In this expression, n represents the number of indistinguishable RACs shared between a pair of unrelated outsoles and m represents an unknown number of distinguishable RAC pairs

that were not evaluated. For each pair of shoes in the database, the number of *indistinguishable* RACs (based on a chosen threshold of indistinguishability) in the same position on both shoes was counted. However, the number of *distinguishable* RACs between these two shoes was not counted or reported, meaning that m is unknown for each shoe in this database. In contrast, Fig. S2 highlights an example using a dataset of three shoes where both n and m can be determined. In these images, an arbitrary grid is utilized to indicate RAC position and simple shapes are used to represent RAC geometries. When comparing Shoe A (S2(a)) to Shoes B (S2(b)) and C (S2(c)) there is one indistinguishable RAC in the same position on all three (the circle in the upper left) and a second indistinguishable RAC in the same position on Shoe C (the box in the upper right). This means that $n = 1$ when compared to Shoe B and $n = 2$ for Shoe C. There are also seven RACs that do not have a corresponding indistinguishable RAC on Shoe B, and six RACs that do not have a corresponding indistinguishable RAC on Shoe C. This means that $m = 7$ for Shoe B and $m = 6$ for Shoe C. Therefore, Shoe A has a RAC-RMF$_{(7|n\,=\,1)}$ with Shoe B and a RAC-RMF$_{(6|n\,=\,2)}$ with Shoe C. Using the same method, Shoe B has a RAC-RMF$_{(4|n\,=\,1)}$ with both Shoe A and Shoe C, and Shoe C has a RAC-RMF$_{(4|n\,=\,2)}$ with Shoe A and a RAC-RMF$_{(5|n\,=\,1)}$ with Shoe B. The general notation for a varied or unknown m when comparing Shoe A to all other shoes would be a RAC-RMF$_{(m|n\,\geq\,1)}$ of 2 out of 2.

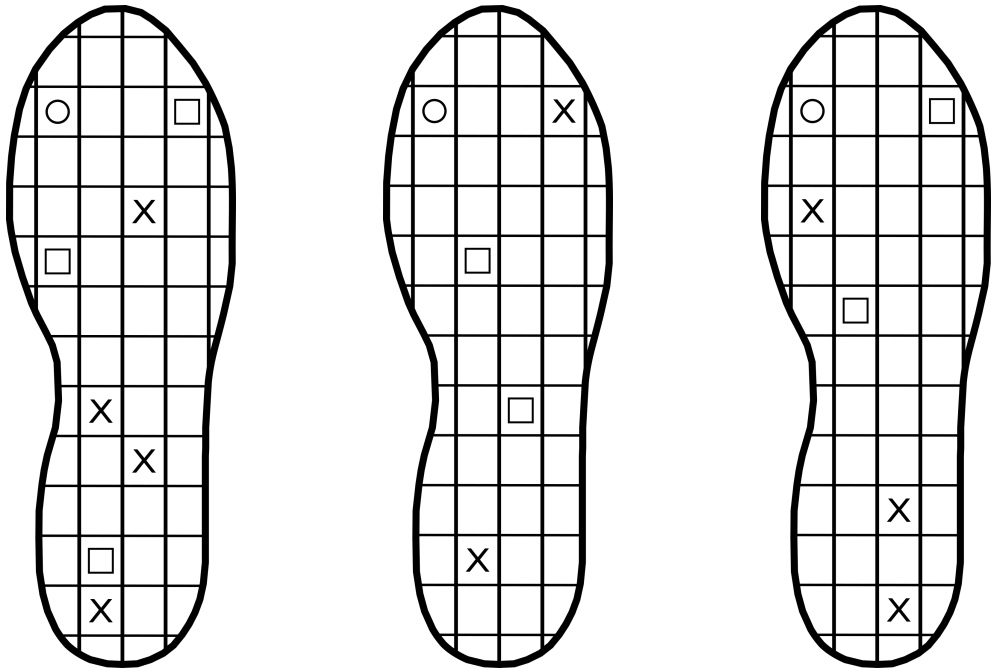


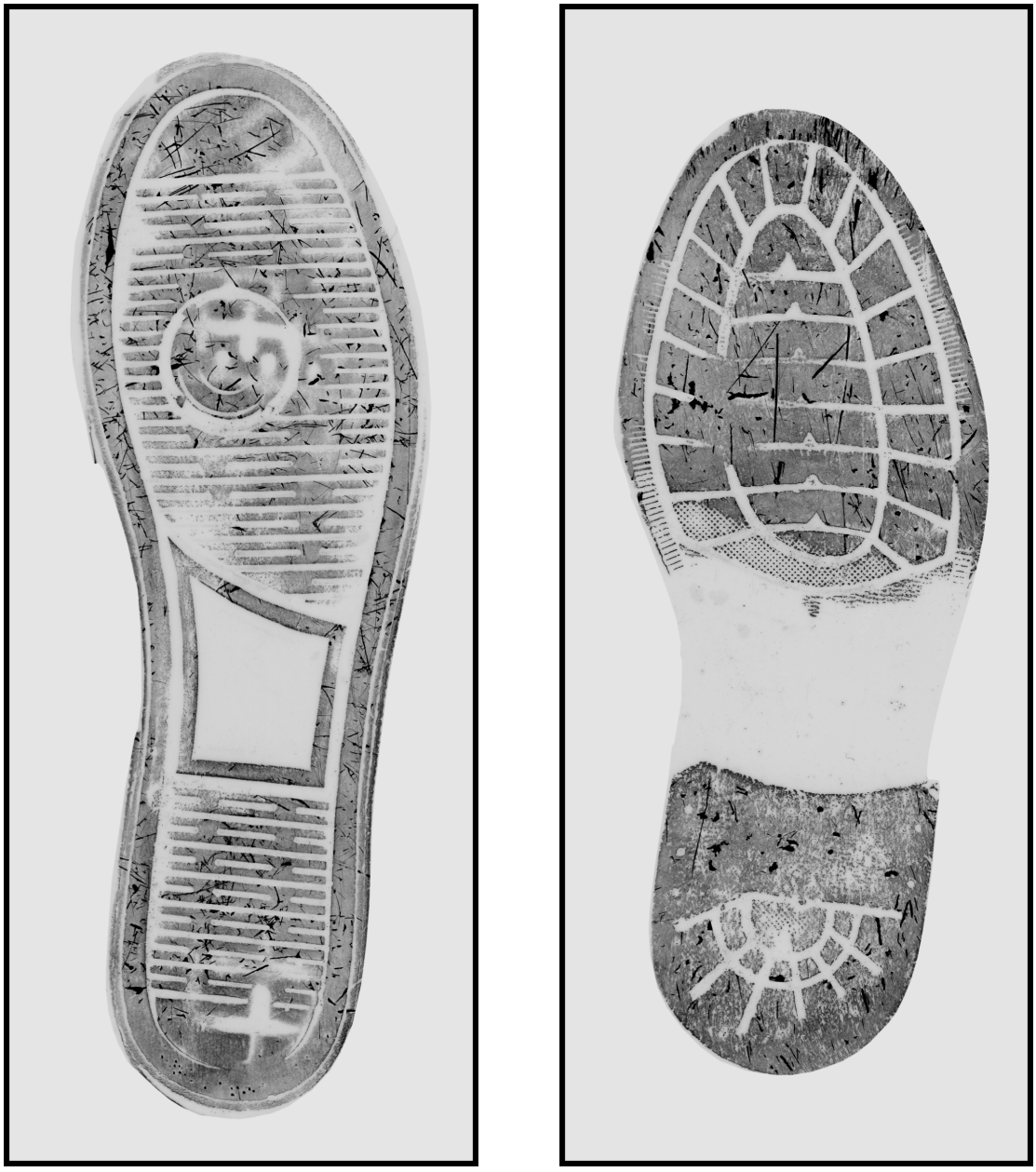
Figure S2: RAC-RMF examples for shoes with known n and m values, where the notation for Shoe A (a) is RAC-RMF$_{(7|n\,=\,1)}$ with Shoe B and RAC-RMF$_{(6|n\,=\,2)}$ with Shoe C.

S3. Updated RAC Counts

The shoes and procedures used to extract randomly acquired characteristics from this dataset are fully described in Speir *et al.* (2016). At the time of publication in 2016, this dataset was comprised of 1,000 outsoles and 57,426 extracted features. Between 2016 and 2019, the dataset grew to 1,300 outsoles and 72,306 extracted features. However, of the 1,300 shoes currently comprising the dataset, 36 were identified (in November 2022) as incompletely extracted. In other words, the number/size of RACs on these outsoles exceeded expectations and as a result, the number of RACs extracted was inadvertently truncated due to a fixed and hard-coded variable in the connected components extraction code. This variable was set based on a compromise between expectation and efficiency. Upon identifying this error of incomplete extraction in a single shoe, all 1,300 shoes were re-evaluated. A total of 36 shoes were determined to have many more RACs than expected, and/or features that are much larger (or geometrically connected) than anticipated. These shoes were re-extracted by increasing the aforementioned fixed variable. Although 36 shoes comprise less than 3% of all outsoles in this dataset, these shoes had a surprisingly large number of marked wear characteristics, which resulted in an additional 8,383 extracted features — or an increase in detected features of nearly 12% (8,383 out of the previously extracted 72,306).

Fig. S3 illustrates two of the 36 shoes that were part of this incompletely extracted set. Close inspection indicates that there are features on these shoes that could be RACs, but remain unmarked, suggesting that the shoes might possess an even larger number of features than originally identified/marked. However, this has not been confirmed; these features may be voids in the fingerprint powder used to create the exemplar, and therefore not visible on the outsole with 4X magnification and oblique illumination. In other words, the shoes have not been remarked a second time, and instead, only re-extracted based on their original mark-up. With this in mind, it is equally important to note that the original 1,000 outsoles were marked by five analysts, over a 15-month time period. As a quality control evaluation, the research group previously reported on the inter-analyst marking similarity between 137 paired RAC maps. Results indicated a correlation of approximately 0.66 based on combined detection and geometric tracing. Further inspection indicated that the majority of variation was in *detection*, such that if two analysts equally detected a feature, the feature was marked relatively consistently over time, making failure to detect the larger source of variation (see Speir *et al.* (2016) pp. 406-407 for additional details).

In summary, the dataset used to report RAC-RMF in this study contains more than 80,600 extracted features, corrected for the 36 shoes previously incompletely extracted, but the procedures used to identify, mark, and extract features (save for a hard-coded variable increasing from 1,000 to 15,000) are the same as previously described, and the intra- and inter-analyst consistency in detection and marking previously reported by Speir *et al.* (2016) for the first 1,000 outsoles remains unchanged.



(a) (b)

Figure S3: Two shoes with more features than originally extracted. The shoe on the left (a) is a Polo Men's casual sneaker (shoe ID 093L with 1,196 marked features of which only 85 were originally extracted). The shoe on the right (b) is a Men's oxford dress shoe but of unknown make/model (shoe ID 094L with 803 marked features of which only 362 were originally extracted).

S4. Shoe Specifications

The distribution of shoe manufacturers for the shoes that shared one or more indistinguishable RAC pairs with an unrelated shoe for at least one of the three thresholds ($t = 1.0$,

$t \geq 0.75$, or $t \geq 0.5$) is shown in Table S1 as both a count and percentage. A total of 890 shoes had at least one indistinguishable pair, but only those manufacturers that comprised more than 2% of the 890 shoes are reported. Nike® is the most common manufacturer for these shoes. However, when considering the entire database of shoes, Nike® is also the most common (40%), followed by Reebok® (12%), and Under Armor® (9%). All other manufacturers comprise less than 3% of the shoes in the database.

Table S1: Distribution of manufacturers for the 890 shoes that shared at least one indistinguishable RAC pair with an unrelated shoe for at least one of the three thresholds ($t = 1.0$, $t \geq 0.75$, or $t \geq 0.5$). Only those manufacturers that comprised more that 2% of the 890 shoes are reported. Registered trademark symbols are omitted for brevity in expression.

Manufacturer	**Count**	**Percent (%)**
Nike	352	39.6
Reebok	87	9.8
Under Armour	78	8.8
Adidas	27	3.0
Asics	24	2.7
Converse	24	2.7
Hoka	19	2.1
New Balance	18	2.0

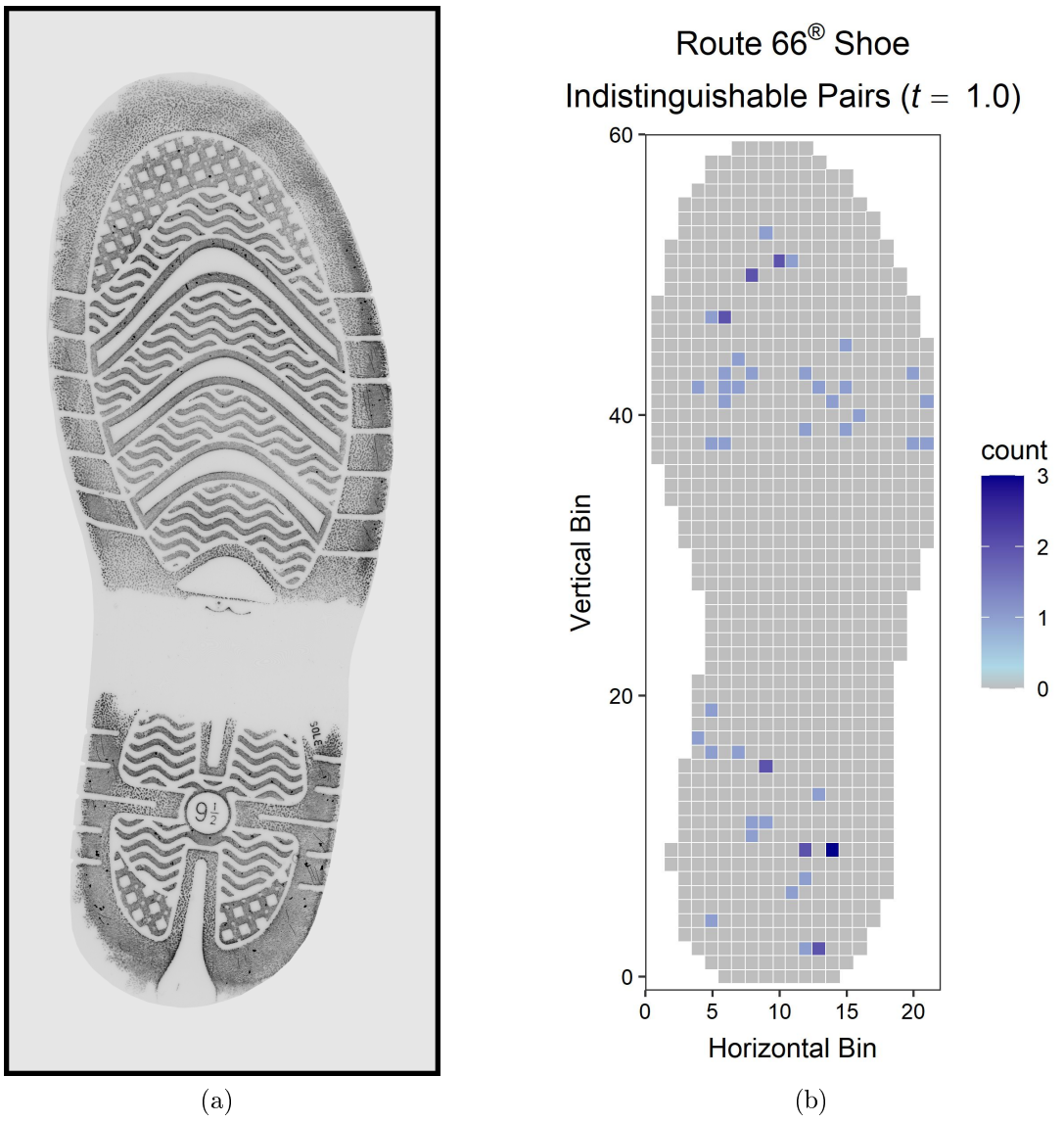
Table S2 shows the most common manufacturer(s) for each of the shared pair groups. If multiple manufacturers had an equal number of shoes in a group, all are shown. Nike® is the most common in the one and two shared pairs group, while Asics® has the majority in the three shared pairs group. The two shoes that share five shared RAC pairs between them are manufactured by Boxfresh® and Polo Ralph Lauren®.

Table S2: Most common manufacturer(s) for each of the shared pair groups. If multiple manufacturers had an equal number of shoes in a group, all are shown. Registered trademark symbols are omitted for brevity in expression.

Shared RAC Pairs	**Manufacturer**	**Count**	**Percent (%)**
1	Nike	352	39.6
2	Nike	23	37.1
3	Asics	2	50.0
4	—	—	—
5	Boxfresh, Polo Ralph Lauren	1	50.0

S5. Shoes with Highest RAC-RMF Values

At a threshold of $t = 1.0$, one shoe had a RAC-RMF$_{(m|n \geq 1)}$ of 49 out of 1,299. This shoe was a Men's Route 66® walking shoe with 358 RACs that shared a total of 51 indistinguishable pairs with 49 unrelated outsoles. Fig. S4(a) shows the marked Handiprint of this shoe, and Fig. S4(b) shows the distribution of the indistinguishable pairs across the standardized outsole. These pairs were distributed across the toe and the heel, with a maximum of three pairs in a cell.

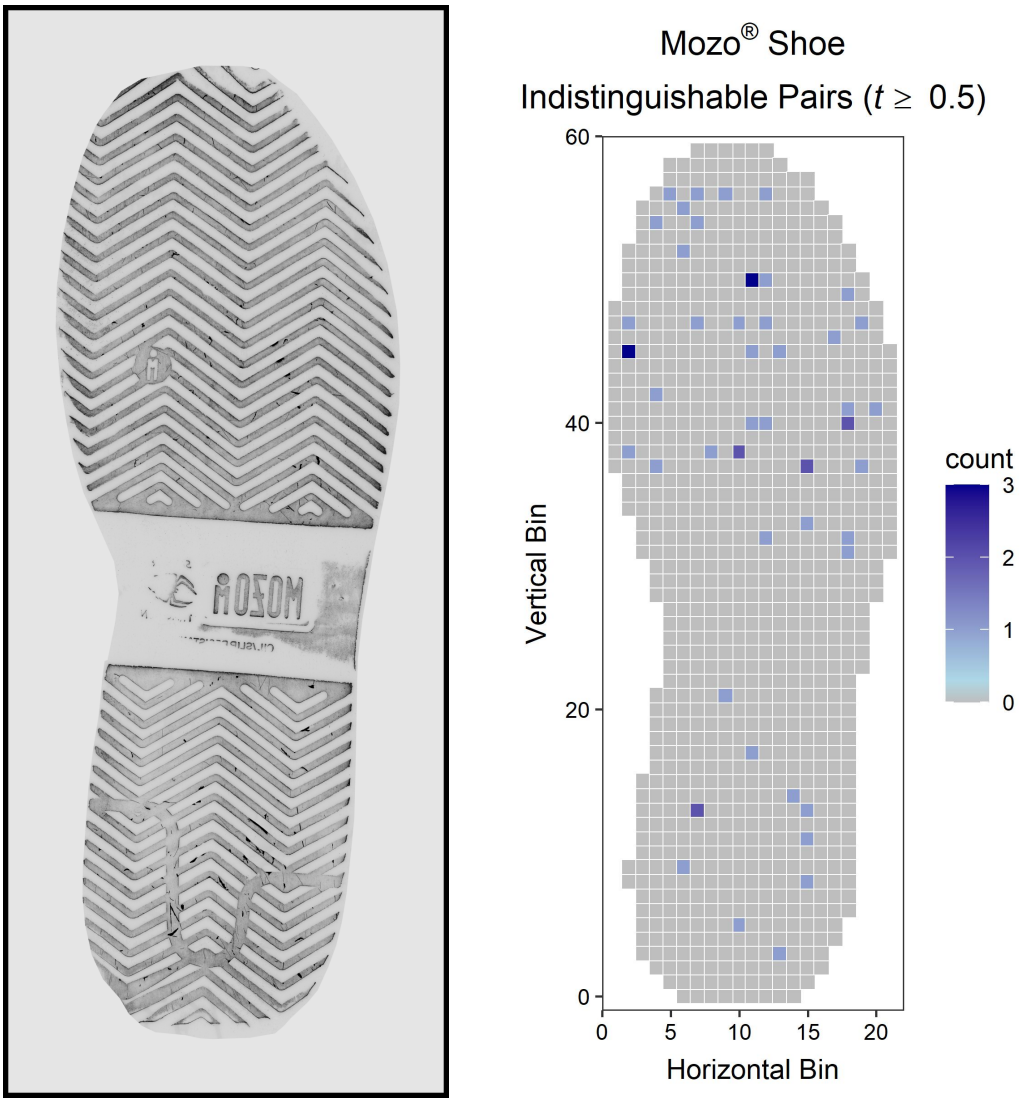


(a) (b)

Figure S4: Marked Handiprint of the Route 66® shoe with 358 RACs (a) and the spatial distribution of the 51 indistinguishable pairs shared with unrelated outsoles (b).

At a threshold of $t \geq 0.5$, a second shoe had a RAC-RMF$_{(m|n \geq 1)}$ of 49 out of 1,299. This shoe was a Men's Mozo® shoe with 702 RACs that shared a total of 54 indistinguishable

pairs with 49 unrelated outsoles. Fig. S5(a) shows the marked Handiprint of this shoe and Fig. S5(b) shows the distribution of the 54 indistinguishable pairs across the standardized outsole. These pairs were slightly more concentrated in the toe than for the Route 66® shoe, but pairs were still distributed across the entire outsole. There was a maximum of three indistinguishable pairs in a cell.



(a) (b)

Figure S5: Marked Handiprint of the Mozo® shoe with 702 RACs (a) and the spatial distribution of the 54 indistinguishable pairs shared with unrelated outsoles (b).

Figs. S6 and S7 show a semi-random sampling of four of the indistinguishable pairs from the Route 66® shoe and the Mozo® shoe, respectively. In all examples, the RAC on the left of each pair of images is the RAC from the Route 66® (S6) or the Mozo® shoe (S7), and the RAC on the right is from an unrelated outsole.

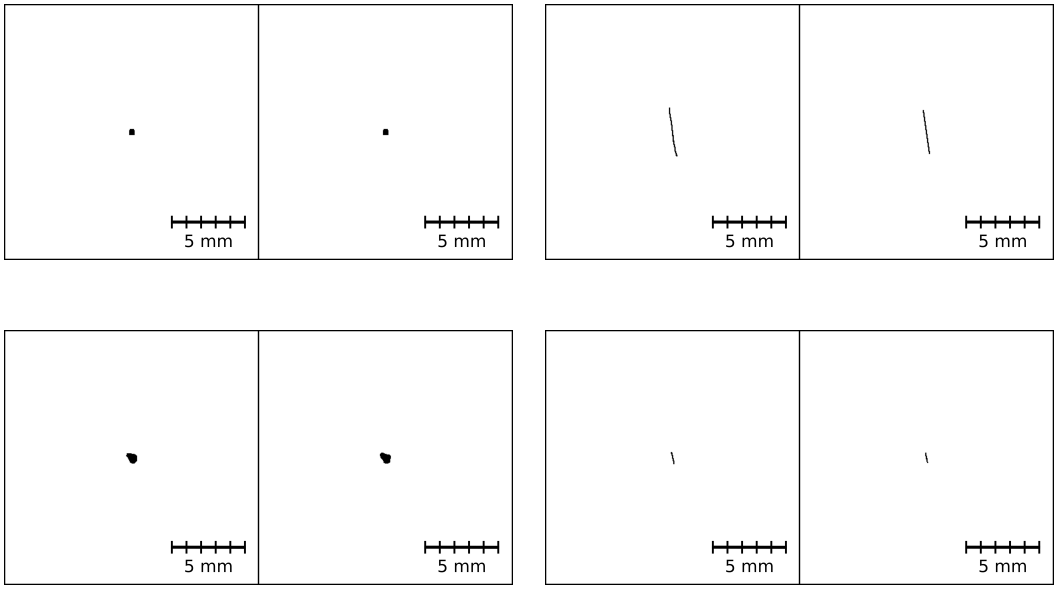


Figure S6: A semi-random sampling of the indistinguishable pairs shared between the Route 66® shoe and unrelated outsoles. The RAC on the left of each image pair is from the Route 66® shoe.

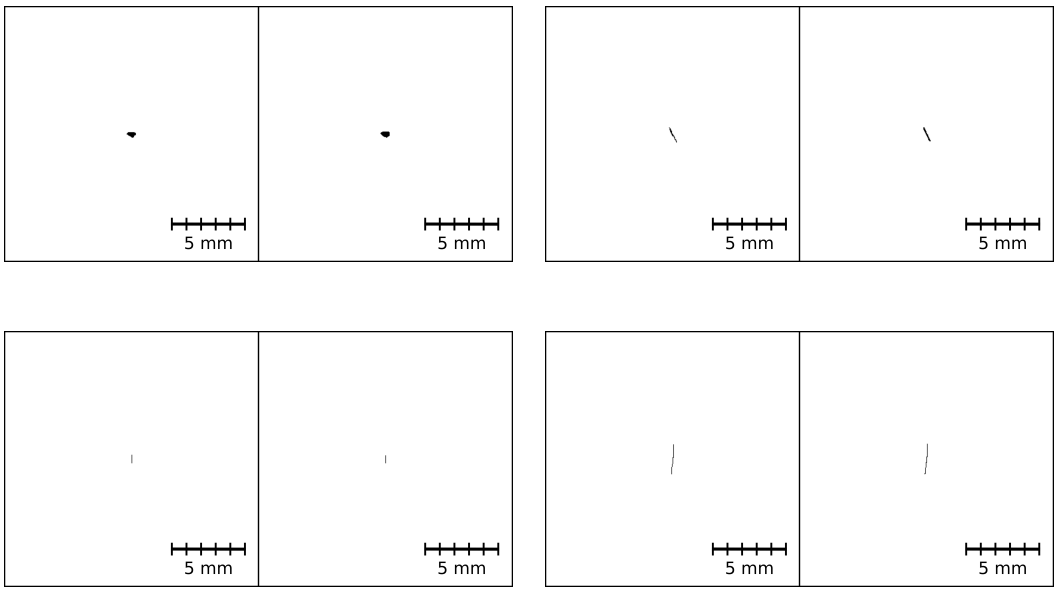


Figure S7: A semi-random sampling of the indistinguishable pairs shared between the Mozo® shoe and unrelated outsoles. The RAC on the left of each image pair is from the Mozo® shoe.

References

* N. Richetelli, W. J. Bodziak, J. A. Speir, Empirically observed and predicted estimates of chance association: Estimating the chance association of randomly acquired characteristics in footwear comparisons, Forensic Science International 302 (2019) 1–14. doi:https://doi.org/10.1016/j.forsciint.2019.05.049.
* J. A. Speir, N. Richetelli, M. Fagert, M. Hite, W. J. Bodziak, Quantifying randomly acquired characteristics on outsoles in terms of shape and position, Forensic Science International 266 (2016) 399–411. doi:https://doi.org/10.1016/j.forsciint.2016.06.012.

3. RAC-RMF of Simulated Crime Scene Impressions in Blood

This manuscript and the associated supplemental material have been submitted for publication: A. N. Smale, J. A. Speir, Estimate of the random match frequency of acquired characteristics in footwear: Part I — Simulated crime scene impressions in blood [Manuscript submitted for publication] (2023).

Estimate of the Random Match Frequency of Acquired Characteristics in Footwear: Part I — Simulated Crime Scene Impressions in Blood

Alyssa N. Smale[a], Jacqueline A. Speir[a,*]

[a]*Forensic & Investigative Science, West Virginia University, 208 Oglebay Hall P.O. Box 6121, Morgantown, WV, 26506, United States*

The aim of this study was to estimate random match frequency of randomly acquired characteristics (RAC-RMF) for simulated crime scene impressions. Part I of this investigation reports this metric using a dataset of more than 160 questioned impressions created in blood and deposited on tile. A total of 759 RACs were identified in the blood impressions and compared to RACs with positional similarity in test impressions from 1,299 unrelated outsoles. Geometric similarity was quantified using a combination of visual comparisons and mathematical modeling based on percent area overlap. Results indicated that RACs in blood impressions were typically smaller, and therefore exhibited a two-thirds increase in the number of indistinguishable pairs compared to their mated test impressions. For shoes contributing at least one RAC, relative RAC-RMF values ≥ 0.0008 were encountered at a rate between 3.4% and 34% for the blood impressions examined in this study. Part II of this investigation provides analogous results based on dust impressions deposited on paper and tile. Although the results in Part I and Part II are specific to randomly acquired characteristics and do not translate into an impression-wide RMF estimate, this research shows that RACs in simulated questioned impressions of the type and quality expected in casework co-occur in position and geometry with RACs in non-mated test impressions. Since theoretical models have traditionally been the basis for estimating RAC-RMF in footwear, the overall contribution of this research to the forensic footwear community is a calibration of this estimate based on empirical data.

Keywords: footwear, randomly acquired characteristics, random match frequency, percent area overlap, simulated crime scene impressions, blood impressions

1. Introduction

Footwear impression evidence is often present at crime scenes, but is much less likely to be collected than other evidence types, such as fingerprints or DNA. This can be attributed

*Corresponding author

Email address: Jacqueline.Speir@mail.wvu.edu (Jacqueline A. Speir)

to the fact that footwear impressions are deposited on the ground in a variety of media, making them easily overlooked and potentially difficult to visualize under normal illumination conditions [1]. Contamination of footwear impression evidence can also result from responding personnel, making it necessary and sometimes time-consuming to differentiate between footwear impressions made during the commission of a crime and those which originated from events that occurred during an emergency response or investigation. However, when identified and properly collected, a footwear impression can provide valuable information about the shoe that created it, such as make, model, size, and characteristics of use.

After an impression is collected from a crime scene, a footwear examiner can be tasked with determining if the impression could have originated from a known shoe. Such a judgment is based on the clarity, quality, and quantity of class characteristics, wear, Shallamach patterns, and randomly acquired characteristics (RACs) present in the questioned impression, and how similar these features are to a test impression created from a known shoe. The opinion reached from this comparison is a function of training and experience, and while examiners might be able to judge the rarity of observed characteristics based on their knowledge, it is desirable to have supporting quantitative analyses. Previous research to evaluate the rarity of RACs has reported that these features rarely repeat on unrelated outsoles, but the majority of this work was performed using datasets of a relatively small sample size [2–6], or as a function of theoretical models [7–9]. In an effort to investigate if this trend persisted within a larger sample of shoes, the RACs on 1,300 shoes in the West Virginia University (WVU) footwear database [10] were analyzed to determine how often similar RAC geometries co-occur in position on unrelated outsoles [11].

The goal of [11] was to estimate the random match frequency of randomly acquired characteristics (RAC-RMF) for a database of high-quality test impressions. The term random match frequency (as opposed to probability) was used to convey that the reported observations were specific to the database that was analyzed, and not intended for extrapolation to any other population. Briefly, test impressions of 1,300 shoes of various makes, models, and sizes were compared to determine if RACs with positional and geometric similarity existed on unrelated outsoles [11]. Positional similarity was determined by mapping the RACs on each shoe to a standardized outsole, where RACs in the same 5 mm $\times$ 5 mm spatial cell were considered to co-occur [10]. Geometric similarity was determined by two different methods: visual assessment and mathematical modeling based on a percent area overlap similarity score [11, 12]. From a total of 80,668 RACs on these 1,300 shoes, RAC-RMF$_{(m|n \geq 1)}$ values as high as 49 out of 1,299 were observed, meaning that a shoe in the database shared at least one indistinguishable RAC pair ($n \geq 1$) with 49 of the 1,299 unrelated shoes, but without regard for the presence of an undetermined number of distinguishable RAC pairs (m) [11].

While similar RAC geometries and RAC-RMFs greater than zero were observed in this study [11], it is necessary to acknowledge that the impressions in the WVU footwear database are Handiprint exemplars that were created in a controlled laboratory setting [10]. These high-quality test impressions reproduced fine details from the outsole, in contrast to what is often expected from impressions found at crime scenes. As a result, the variation in RAC-RMF when presented with lower-quality impressions more representative of crime scene impressions remains unknown, thus prompting further investigation.

To generate impressions of lower quality, a dataset of simulated crime scene impressions was created using blood as the impression medium. These impressions were examined for RACs and compared to test impressions of unrelated outsoles, allowing for estimation of the RAC-RMF for each shoe in the simulated crime scene impression dataset. Two main factors were anticipated to affect the resulting RAC-RMF values relative to those observed when analyzing high-quality test impressions. First, it was expected that only a portion of the RACs from the outsoles would reproduce in the blood impressions as a result of the medium, substrate, impression clarity, impression totality, RAC size/complexity, or a combination of these variables. With less RACs present in the blood impressions to compare to RACs on unrelated outsoles, the overall number of RAC comparisons would decrease, which in turn would likely decrease the frequency of both low and high RAC-RMF values, while increasing the uncertainty in overall estimates. Second, the influence of the medium and substrate was expected to cause changes in RAC size and shape, which would affect the overall degree of geometric similarity between non-mated RACs. However, the effect of these size and shape changes on RAC-RMF was more difficult to predict, as these differences could make a RAC less similar to some RACs while making it more similar to others.

In light of these hypotheses, it was difficult to predict how the combination of factors might interact, thereby motivating this study. Although the following results are based on simulated crime scene impressions, the estimates of RAC-RMF reported in this study should not be directly translated into casework when presented with blood impressions. Instead, the trends observed when comparing the results of studies with impressions of different type and quality should be used to reinforce and/or update an examiner's internal-calibration. The analysis of both test impressions [11] and simulated crime scene impressions highlights the influence of medium and substrate on the proportion of RACs from the outsole expected to transfer to an impression, the degree of geometric similarity that is possible between unrelated RACs, and the possible chance of erroneous source association based on indistinguishable RAC pairs shared between non-mated outsoles. In an effort to further understand the influence of various media, substrates, and collection techniques on possible RAC-RMF estimates, Part I of this investigation focuses on simulated crime scene impressions created from blood, while Part II will compare and contrast the results of this study with those obtained when examining impressions created in dust.

2. Materials and Methods

A subset of shoes from the WVU footwear database [10] was used to create simulated crime scene impressions in blood. For brevity, the simulated crime scene impressions in this dataset will be referred to as blood impressions. It was anticipated that there would be loss of outsole detail in the blood impressions, so shoes with a large number of RACs were chosen to increase the chance of transferring and identifying RACs in the impressions. The resulting questioned impressions were processed, including chemical enhancement, registration, and RAC identification. This was followed by RAC extraction and localization, ultimately allowing for random match frequency estimates of randomly acquired characteristics (RAC-RMF) in a similar manner as described for high-quality test impressions in [11].

2.1. Simulated Crime Scene Impressions in Blood

To create the blood impressions, a paper towel soaked with defibrinated horse blood (10052-754 Hardy Diagnostics) was placed on top of an acetate sheet and a layer of dry paper towels in the bottom of an aluminum tray. A participant fitting the size of each shoe was chosen to wear the shoe and create a walking impression. The wearer stepped into the tray containing the blood, and then onto a ceramic or vinyl tile to deposit an impression. The four types of tiles used were all purchased from Lowe's® and are shown in Fig. 1 (1(a) cream ceramic (1599903), 1(b) striped tan vinyl (737993), 1(c) patterned tan vinyl (1346018), and 1(d) white vinyl (1346019)). A total of 165 blood impressions were created. After a minimum of 72 hours, the dried impressions were scanned at a resolution of 600 pixels per inch (PPI) using an Epson® Expression® 11000XL Graphic Arts scanner as described in [13].

(a) (b) (c) (d)

Figure 1: Examples of blood impressions on four different tile types: (a) cream ceramic, (b) striped tan vinyl, (c) patterned tan vinyl, and (d) white vinyl.

Leucocrystal violet (LCV) was used to enhance both patent and latent features on the tiles. A solution was prepared as outlined in [1] using the following reagents: leucocrystal violet (AC204330250 Acros Organics), sodium acetate (216570-250G Beantown Chemical), 5-sulfosalicylic acid (124400-100G Beantown Chemical), and hydrogen peroxide (470301-282 Ward's Science). The prepared LCV solution was sprayed directly onto the impression. In the presence of hemoglobin, hydrogen peroxide served as a catalyst for the oxidation of LCV to crystal violet, which caused the areas of the tile containing blood to turn a violet color [1]. After approximately 30 seconds, a second spray bottle was used to rinse the tile thoroughly

with water. Each tile was allowed to dry for a minimum of 24 hours in a vertical position, and then the chemically-enhanced impression was scanned at a resolution of 600 PPI using the same Epson® scanner [13]. To illustrate, Fig. 2 shows scans of the impressions in Fig. 1 following enhancement with leucocrystal violet. Only the enhanced impressions were used for analysis for the remainder of the study.

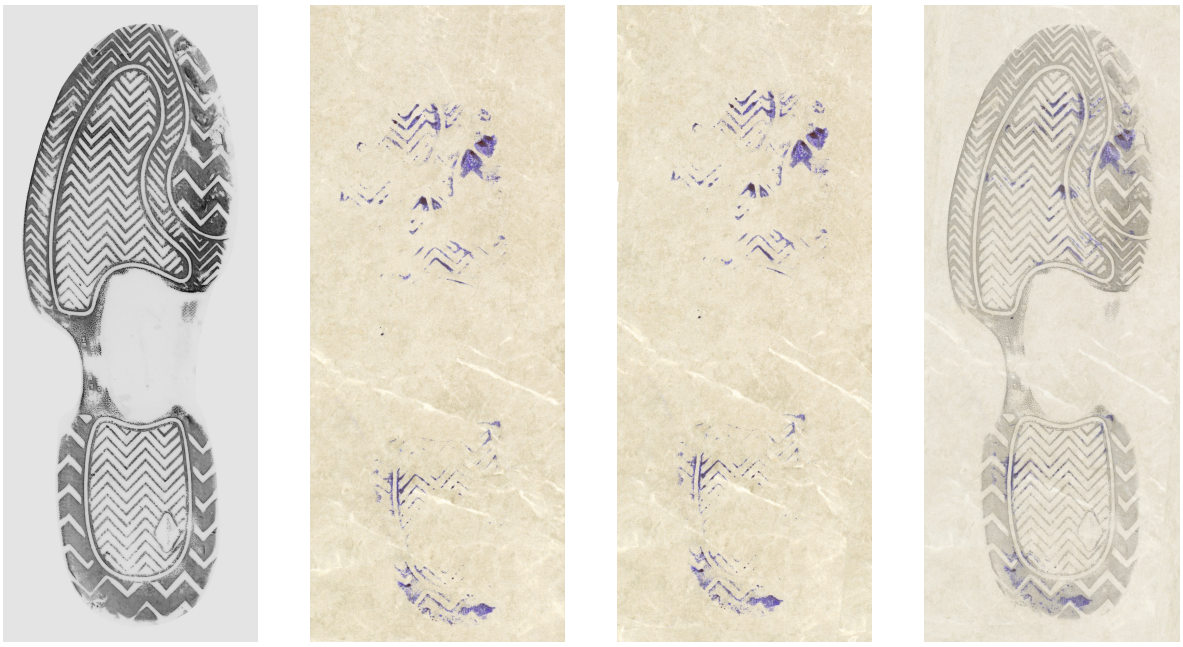
(a) (b) (c) (d)

Figure 2: Examples of blood impressions from Fig. 1 enhanced with leucocrystal violet on four different tile types: (a) cream ceramic, (b) striped tan vinyl, (c) patterned tan vinyl, and (d) white vinyl.

2.2. Post-Processing

The resulting chemically-enhanced images were registered to a scanned Handiprint of the corresponding outsole [10] using one of two methods. The first method, previously outlined by Speir *et al.* [10], required the selection of eight ground control points in both images and used a warping function to perform the registration. This method was used when it was possible to reliably select the same features in both images. If this could not be successfully accomplished, the second method was performed in Adobe® Photoshop® where the blood impression was manually translated, rotated, and scaled, as necessary, to achieve alignment with the test impression. There were three impressions that could not be adequately registered to the corresponding Handiprint using either method due to distortion, reducing the total number of blood impressions to 162. Of those that were successfully registered, 37 were automatically aligned using ground control points and 125 were aligned manually using Adobe® Photoshop®. Results typical of both types of alignment are illustrated in Fig. 3, including the Eigen registered Handiprint test impression [10] (3(a)), the pre-registered blood impression (3(b)), the final registered blood impression (3(c)), and an overlay of the

test and blood impressions after alignment (3(d)), which was inspected as a quality control step.



(a) (b) (c) (d)

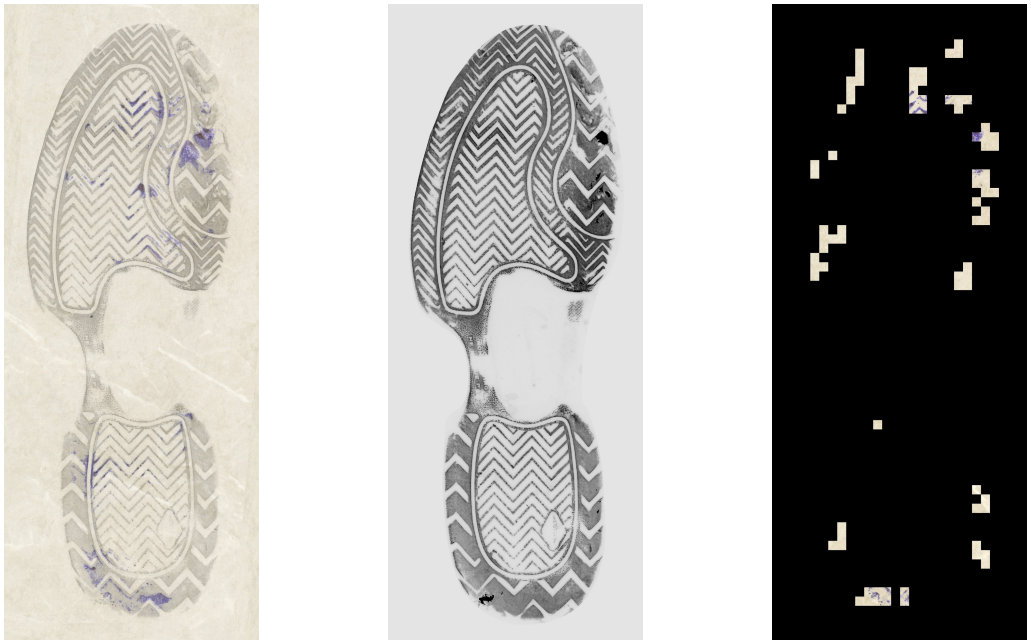
Figure 3: Results typical of both registration procedures for aligning a blood impression (b) to its corresponding Eigen registered Handiprint test impression [10] (a). The final registered version of the blood impression is shown in (c), and an overlay of the two impressions after alignment is shown in (d).

2.3. Marking of RACs

When marking RACs in the blood impressions for this study, an ideal approach would utilize linear sequential unmasking, where the questioned impression is analyzed completely prior to viewing the exemplar impression [14]. However, in an attempt to identify as many authentic RACs as possible without marking pseudo-accidentals (*e.g.*, voids in the medium due to lack of saturation on the outsole or lack of transfer during contact), a modified sequential unmasking approach was implemented using a two-step procedure. In the first step, the marked test impression [10] corresponding to each blood impression was used to create a "patch" map. This map consisted of a combination of opaque and transparent patches (each approximately 4.7 mm $\times$ 4.7 mm in size). When overlaid on the corresponding blood impression, the opaque patches denoted locations where RACs were known to be absent, and therefore not useful for further inspection (since a suspected RAC in these locations would ultimately be deemed an erroneous mark-up or a pseudo-accidental). Conversely, transparent patches represented physical locations where one or more RACs were confirmed to exist in the mated test impression. An example is shown in Fig. 4 using the registered questioned impression pictured in Figs. 3(c) and 3(d). After alignment of the test and blood impressions (Fig. 4(a)), the previously-marked test impression [10] (Fig. 4(b)) was used to

create a patch map that highlighted areas where RACs could be expected in the blood impression. When overlaid onto the blood impression (Fig. 4(c)), the patch map was used to direct the search for RACs without bias in terms of the number of RACs found in each patch, the exact position of the RACs within the patch, or the size and shape of the RACs. Suspected RACs were traced by a researcher in Adobe® Photoshop® using the pencil tool (width = 2 pixels) and filled using the bucket tool when necessary, as described in [10].

In the second step of the modified sequential unmasking approach, the marked blood impression was overlaid on the previously-marked test impression. Using the identified RACs from the high-quality test impression, four outcomes were possible. First, the RACs marked in the blood impression were confirmed. Second, RACs marked in the blood impression were deemed erroneous or pseudo-accidentals and were removed. Third, the shapes of RACs marked in the blood impression were adjusted. Fourth, RACs were newly-marked if missed in the first step. The purpose of this second step was two-fold; to ensure that all RACs that transferred to the blood impression were identified, and to avoid marking pseudo-accidentals. A negative consequence of this step was that any RAC marked or updated after unmasking was geometrically-biased due to *a priori* knowledge of the RAC's shape in the test impression. To investigate the effect of this bias on shape, each RAC was tagged with a label to denote during which step it was marked, making it possible to differentiate RACs identified prior to viewing the test impression (unbiased) from those identified after (biased).



(a) (b) (c)

Figure 4: The overlay of the test and blood impressions after registration (a) is reproduced from Fig. 3. The marked test impression [10] (b) was used to create a patch map that was overlaid on the corresponding registered blood impression (c) to focus the researcher's efforts when attempting to locate RACs in the questioned impression.

After completing both steps, all marked features were extracted from the blood impression to create a RAC map showing the shape and location of identified RACs. Although each blood impression was registered to its corresponding Handiprint, slight mis-registration or a shift in the RAC's centroid could cause a RAC in the blood impression RAC map to be localized to a different 5 mm $\times$ 5 mm spatial cell on the normalized outsole [10] than the same RAC extracted from the test impression. To account for this, the spatial cell of the RAC in the test impression was used as ground truth to determine the appropriate spatial cell for the RAC from the blood impression. Lastly, each RAC from the blood impressions was classified as either linear, compact, or variable in shape [12].

2.4. Quality Control

When considering both the test impression and blood impression of the same shoe, it was expected that there would be variation in the shape and number of the RACs observed. This variation can result from medium and substrate interference, the dynamic walking method used to create the blood impressions, and/or the marking ability of the researchers who marked the test and/or blood impressions. Of these, a quality control step was implemented to determine the impact of researcher marking. The variability associated with this step could include failing to identify RACs in an impression (or differences in a researcher's ability to detect RACs), and/or variations in the manner in which RACs are marked (length, width, etc.), as well as bias in RAC marking based on simultaneously viewing an unmasked exemplar and questioned impression. Each scenario is further described below, except for the rate of RAC identification. In other words, RAC identification is an analyst/researcher/examiner-specific attribute based on past experience and training, and although quantified in this study, was not directly investigated.

2.4.1. Repeatability in Marking

A random selection of 17 blood impressions, or approximately 10% of all impressions, was copied and mixed into the dataset prior to marking. This resulted in a second blind marking of each of these chosen impressions by the researcher. Although the RAC counts for the duplicate markings were compared, the researcher's repeatability in detecting RACs was not under investigation. Instead, the purpose of duplicate marking was to evaluate changes in RAC shape as marked by the researcher. This was accomplished by comparing RACs that were detected in both markings using percent area overlap to determine the degree of similarity, and the minimum, maximum, average, and standard deviation of this metric. The average difference in percent area overlap from 100% was used to estimate variation, and this difference was propagated to determine its impact on the uncertainty in probability of indistinguishability predictions.

2.4.2. Bias in Marking

The effect of researcher bias on RAC shape was investigated by comparing each RAC from the blood impressions to its known mated RAC from the corresponding test impression. These RAC pairs were separated into two groups based on whether the RAC from the blood impression was identified during the first step of marking (unbiased shape) or the

second step of marking after seeing the known mate in the test impression (biased shape). Potential differences in both percent area overlap value and probability predictions from the mathematical model between these groups were evaluated using a chi-square test of independence [15] on binned results to determine if the effect of bias during marking was statistically significant.

2.5. RAC Comparisons

The flowchart depicted in Fig. 5 is an outline of the different RAC comparisons that were performed for this study and in previous related studies [10–12]. The arrow colors will be referenced moving forward to designate a specific series of steps. Note that some results and abbreviations which have not yet been discussed are included in this figure, but will be further explained in subsequent sections. This outline highlights the relevant RAC comparisons between known mated (KM) and known non-mated (KNM) RAC pairs, as well as provides the appropriate reference for results from past studies.

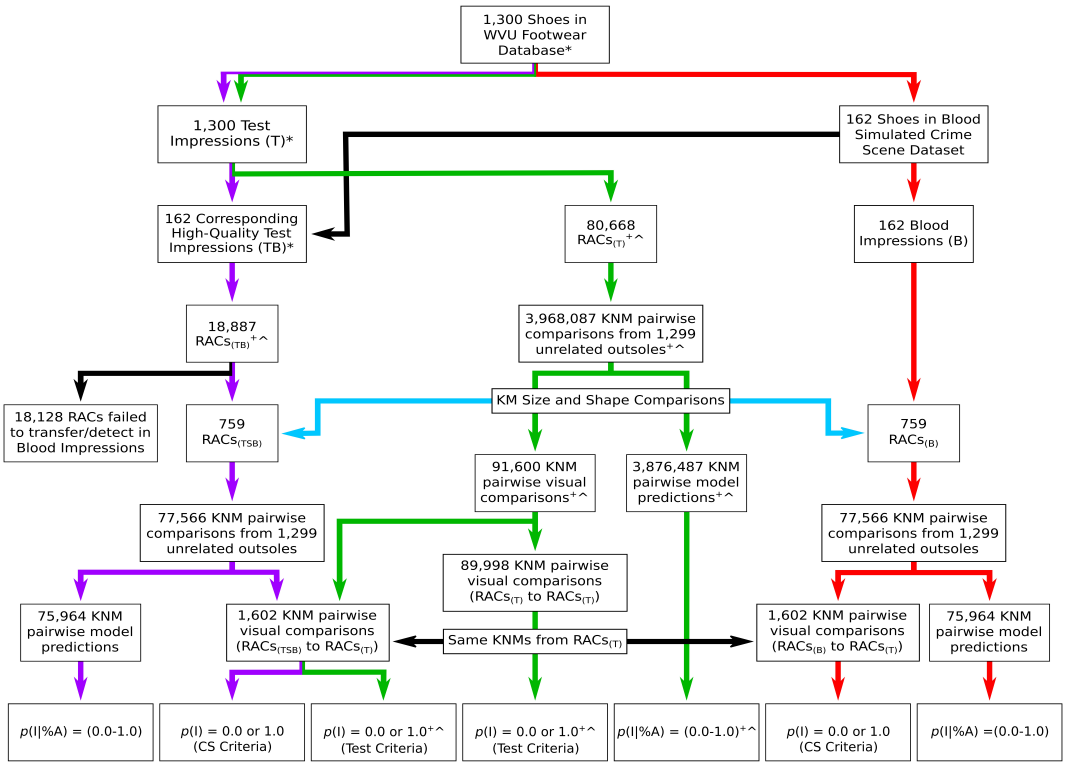


Figure 5: Outline of the RAC comparisons performed in this study and in previous work (*Speir *et al.* (2016) [10], $^{+}$Richetelli *et al.* (2019) [12], ^Smale and Speir (2023) [11]). Please refer to this figure when reading the following sections.

2.5.1. Known Mated RAC Pairs

The mated RAC from the corresponding test impression was identified for each RAC that reproduced in a blood impression, comprising a test impression subset (TSB). Each RAC from the blood impressions ($\mathrm{RAC}_{(\mathrm{B})}$) was compared to its mated RAC in the corresponding

test impression ($RAC_{(TSB)}$) to investigate any variations in RAC size and shape between impression types as a result of medium and substrate influences (Fig. 5, blue arrow). To assess changes in RAC size, the area of the same RAC in both impressions was computed. To assess changes in RAC shape, the category of each RAC in a mated pair was compared to determine if any variations in RAC shape resulted in a different classification.

2.5.2. Known Non-Mated RAC Pairs

In the previous study of RAC-RMF using test impressions [11], all RACs from test impressions ($RACs_{(T)}$) were compared to $RACs_{(T)}$ with positional similarity in test impressions made by unrelated outsoles (Fig. 5, green arrows). After marking all $RACs_{(B)}$ in the blood impressions and identifying their known mated $RACs_{(TSB)}$ in the corresponding test impressions, two additional sets of comparisons between non-mated RAC pairs were possible. The first set compared $RACs_{(B)}$ in the blood impressions versus non-mated $RACs_{(T)}$ in test impressions made by 1,299 unrelated outsoles in the full database (Fig. 5, red arrows). The second set compared $RACs_{(TSB)}$ from the test impressions which reproduced in a blood impression (referred to as the test impression subset) versus the same non-mated $RACs_{(T)}$ in test impressions made by the same 1,299 unrelated outsoles (Fig. 5, purple arrows).

RAC pairs from unrelated outsoles were compared using a combination of visual comparisons and predictions from a mathematical model based on a percent area overlap (%A) similarity score [12]. Of the 91,600 visual comparisons previously performed [12] (Fig. 5, green arrows), 1,602 contained at least one of the $RACs_{(TSB)}$ in the test impression subset. Note that for these comparisons, RAC pairs were deemed indistinguishable by only allowing for minor variation expected in replicate test impressions [12] (referred to as "Test Criteria" in Fig. 5). To account for different researchers and the passage of time since the original visual comparisons were performed, these 1,602 RAC pairs were visually-compared once again by two researchers (Fig. 5, purple arrows). In addition, an updated criteria was used to determine indistinguishability for these pairs by allowing for slightly greater variation that could be expected when comparing crime scene and test impressions (referred to as "CS Criteria" in Fig. 5). The known mated $RACs_{(B)}$ from the blood impressions were then substituted into the appropriate pair in order to visually evaluate the similarity between these $RACs_{(B)}$ and the same selection of known non-mated $RACs_{(T)}$ from test impressions (Fig. 5, red arrows). For each RAC pair that was visually-compared, a probability of indistinguishability ($p(I)$) of 0.0 was assigned to pairs deemed distinguishable and a probability of 1.0 was assigned to those determined to be indistinguishable. For RAC pairs not visually-compared, the mathematical model outlined in [12] was re-used to allow for direct comparisons between the results presented in [11] and the two new datasets under investigation in this study. Using this model, the probability of indistinguishability based on the pair's percent area overlap value ($p(I|\%A)$) was predicted for each pair as a value between 0.0 and 1.0.

2.6. Random Match Frequency of Randomly Acquired Characteristics

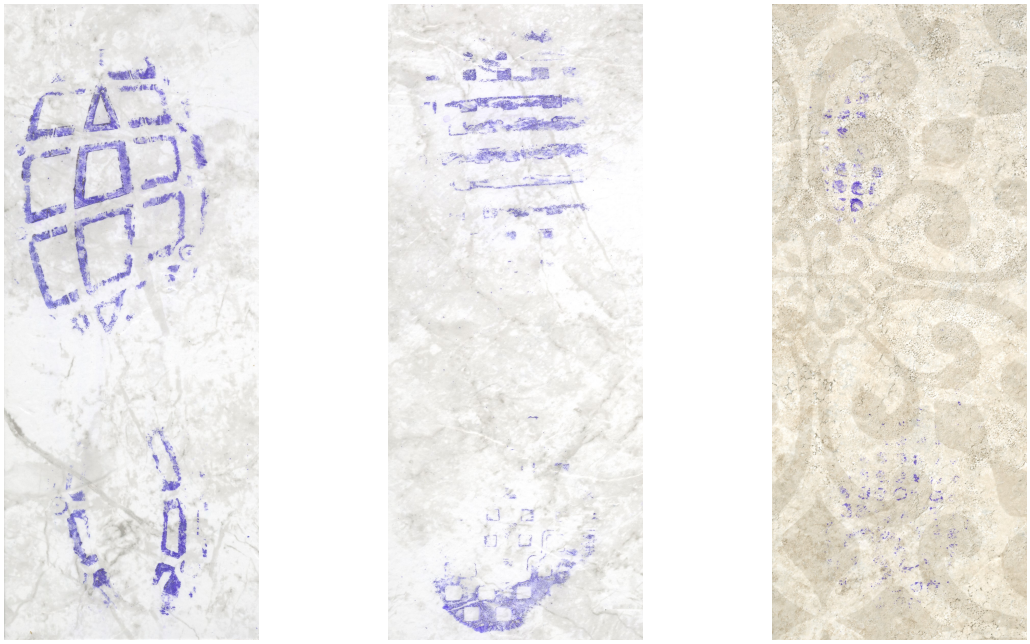
After obtaining $p(I)$ or $p(I|\%A)$ for all RAC pairs as illustrated in Fig. 5, the same procedure as previously described in [11] was performed to compute random match frequency

of RACs — or RAC-RMF$_{(m|n)}$. In this notation, n represents the number of indistinguishable RACs shared between a pair of unrelated outsoles and m represents an undetermined number of distinguishable RAC pairs (please see [11] for additional information). RAC$_{(B)}$-RMF$_{(m|n \geq 1)}$ was computed for each shoe in the simulated crime scene dataset that reproduced at least one RAC in its associated blood impression in order to determine how many test impressions from unrelated shoes shared *one or more* indistinguishable RAC pairs with each blood impression. This process was repeated for the test impression subset to obtain RAC$_{(TSB)}$-RMF$_{(m|n \geq 1)}$. Each pair of unrelated shoes with at least one shared indistinguishable pair ($n \geq 1$) was further evaluated to determine if there were any additional shared pairs in order to report the maximum value of n for these comparisons.

3. Results

3.1. Simulated Crime Scene Impressions in Blood

A total of 165 impressions were created on four different types of tiles by participants wearing a properly-fitted shoe. The use of leucocrystal violet as an enhancement technique worked well to visualize otherwise latent details. In most cases, portions of the outsole did not fully reproduce in the impressions. Following enhancement, there were 58 impressions deemed near-complete, 59 that had a moderate amount of detail (about half of the impression visible), and 48 that had a minimal number of features visible. Fig. 6 shows three impressions of varying degrees of totality, ranging from near-complete (6(a)) to minimal (6(c)).



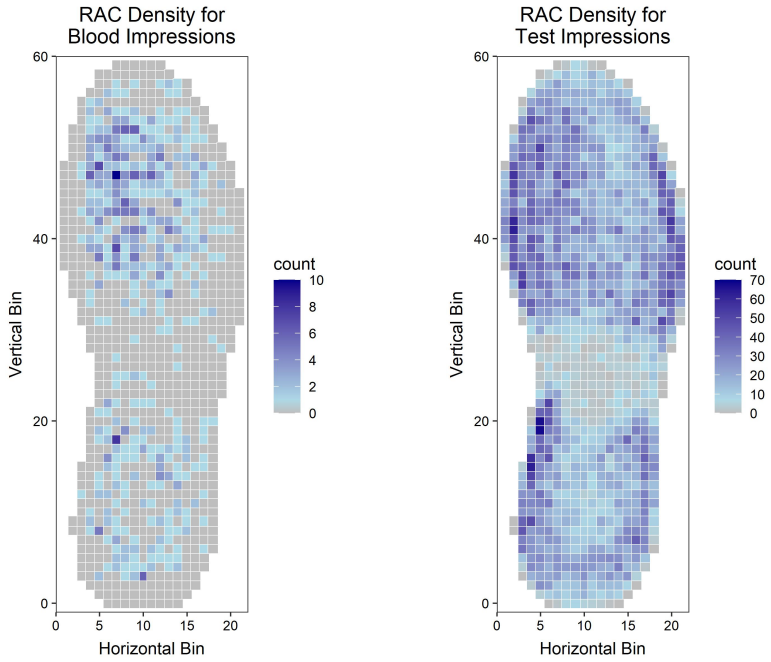
(a) (b) (c)

Figure 6: Blood impressions enhanced with leucocrystal violet displaying varying degrees of totality, with examples of near-complete (a), moderate (b), and minimal (c).

3.2. Marking of RACs

Each of the 162 impressions in the dataset that was successfully registered was marked by a single researcher, and 122 of the impressions (75%) had at least one RAC. There were 353 potential RACs identified using the patch maps. Of these, 105 were deemed valid, 230 were determined to be pseudo-accidentals as a function of the combined medium and substrate, and 18 were confirmed but required an adjustment to the marked shape after viewing the corresponding RAC in the test impression. An additional 636 RACs were identified after unmasking when the blood and previously marked high-quality test impression were simultaneously compared. This resulted in 105 unbiased and 654 biased (636 additional + 18 shape changes) RAC shapes, for a total of 759 RACs (or $\text{RACs}_{(\text{B})}$ from Fig. 5). Note that a total of 782 RACs were originally confirmed at the conclusion of the second marking, but 23 were removed due to either multiple RACs from the blood impression pairing with a single RAC from the test impression, or vice versa (please see supplemental section S1 and Figs. S1 and S2 for additional information and examples).

Of the 759 verified $\text{RACs}_{(\text{B})}$, 330 were classified as variable, 228 were linear, and 201 were compact. Each $\text{RAC}_{(\text{B})}$ was localized to a spatial cell on the normalized outsole based on the known location of its mated $\text{RAC}_{(\text{TSB})}$ from the high-quality test impression (please see supplemental material section S2 and Fig. S3 for additional information). These 759 $\text{RACs}_{(\text{B})}$ were distributed across the outsole in 394 different spatial cells (out of 987), as shown in Fig. 7(a). For comparison, the 162 corresponding high-quality test impressions (TB) had 18,887 $\text{RACs}_{(\text{TB})}$, which were distributed in 958 cells as shown in Fig. 7(b). Note the difference in the maximum value of the density between the two images, as the maximum RAC count in a single cell was 10 for the blood impressions and 68 for the test impressions. Despite the variation in total count and count per cell, RACs were observed to cluster in the medial toe for both impression types. In addition, it can be seen that RACs around the perimeter of the outsole that were visible in the test impressions often failed to reproduce in the blood impressions. However, this is hypothesized to be the result of the different methods used to create the impressions, as the test impressions were created using a static benchtop method [10] which ensured that all portions of the outsole made contact with the Handiprint material, while the blood impressions were created using a dynamic walking method [1] which could have resulted in less contact between the outsole and medium/substrate around the perimeter.



(a) (b)

Figure 7: Spatial distribution of RACs across a normalized outsole [10] extracted from the 162 shoes comprising the blood impressions (a) and their corresponding Handiprint test impressions (b). Note that results have been plotted using two different density scales.

Fig. 8 shows the number of RACs present in the high-quality test impression versus the number that reproduced in the blood impression for each of the 162 shoes in the dataset. The minimum RAC count in a single blood impression was 0, the maximum was 30, and the average was 5, with a 95% confidence interval (CI) of [2, 12] assuming a Poisson distribution [16]. For the corresponding test impressions, the minimum RAC count in a single impression was 31, the maximum was 506, and the average was 117, 95% CI [97, 140]. As expected, the number of RACs that reproduced in the blood impression is not only a function of the number of RACs present on the outsole. The maximum percentage of RACs from a test impression detected in the corresponding blood impression was 21.5% (or 14 transferred out of 65 RACs present). These results demonstrate that although one or more RACs were detected in 75% of the blood impressions, over 95% of RACs did not transfer (only 759 confirmed out of a possible 18,887) based on the quality, clarity, and totality of the impressions in this dataset.

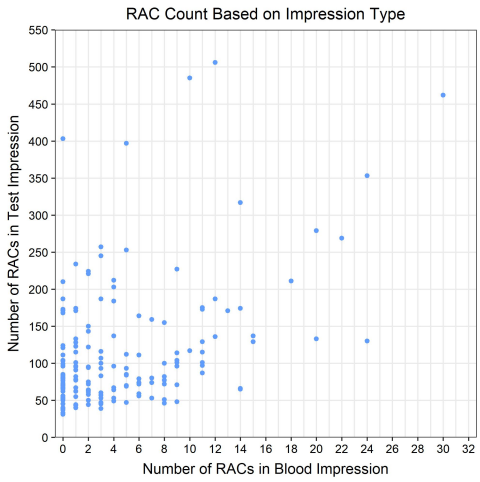


Figure 8: Number of RACs identified in the blood impression versus the number of RACs previously identified in the corresponding high-quality test impression for each of the 162 shoes in the simulated crime scene impression dataset.

In an attempt to determine the actual "size" of the RACs contributing to each dataset, the longest axis of each RAC was extracted, and converted from pixel space to millimeters based on the scanning resolution of 600 PPI. The count and percentage of RACs within different size ranges along with the maximum RAC length are reported in Table 1 for each dataset of interest. The summary data indicates that more than 75% of all identified RACs had a major axis no greater than about 3 mm in length. The area was also computed for these RACs, and the distribution of areas from the 18,887 $\text{RACs}_{(\text{TB})}$ in the 162 test impressions and the 759 $\text{RACs}_{(\text{TSB})}$ that reproduced in a blood impression is shown in Fig. 9, where the x-axis indicates the inclusive upper limit of each bin (*e.g.*, the first set of bars includes RACs with areas (mm^2) in the range (0.0-0.2] and the second set includes (0.2-0.4]). This plot demonstrates that, for the RACs available on these 162 shoes, larger RACs reproduced and were detected in blood impressions more often than smaller RACs.

Table 1: Distribution of RAC sizes as determined by the longest axis for all 18,887 $\text{RACs}_{(\text{TB})}$ in the 162 corresponding test impressions, the 759 $\text{RACs}_{(\text{TSB})}$ in the test impression subset, and the 759 $\text{RACs}_{(\text{B})}$ in the blood impressions. Count is reported first, followed by the percentage in parentheses.

	Count and Percentage (%) of RACs						**Maximum**
	(0–1] mm	**(1–2] mm**	**(2–3] mm**	**(3–4] mm**	**(4–5] mm**	**>5 mm**	**Length (mm)**
18,887 $\text{RACs}_{(\text{TB})}$	5,792 (30.7)	5,875 (31.1)	3,036 (16.1)	1,660 (8.8)	987 (5.2)	1,537 (8.1)	42.2
759 $\text{RACs}_{(\text{TSB})}$	139 (18.3)	228 (30.0)	133 (17.5)	90 (11.9)	78 (10.3)	91 (12.0)	25.2
759 $\text{RACs}_{(\text{B})}$	208 (27.4)	307 (40.4)	125 (16.5)	62 (8.2)	35 (4.6)	22 (2.9)	9.1

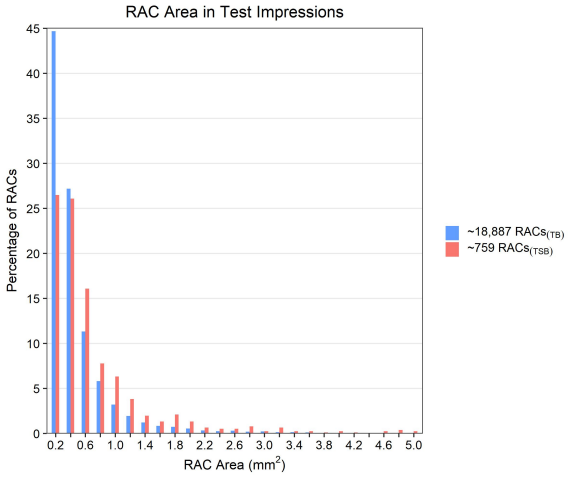


Figure 9: Distribution of RAC areas in test impressions from the 162 shoes used to create blood impressions. The blue distribution reports the area of ~18,887 $\text{RACs}_{(\text{TB})}$, while the red distribution reports the area of only those ~759 $\text{RACs}_{(\text{TSB})}$ which reproduced in a blood impression. Only RACs with an area less than or equal to 5.0 mm^2 are plotted, but this includes approximately 99% of all RACs under investigation.

3.3. Known Mated RAC Pairs

The known mated RAC pairs were compared to investigate changes in RAC size and shape due to impression type (Fig. 5, blue arrow). Fig. 10 shows the relative area of 717 of the 759 RACs in the test versus the blood impressions. Data points above the diagonal line (66%) represent RACs in the blood impressions that were smaller than their test impression mate, while those below the line (34%) represent RACs that were larger in the blood impression. Only three RACs (<0.5%) were the same size in both impressions. This demonstrates that, for the experimental conditions of this study (selected shoes, blood, tiles, etc.), the size of a RAC tends to decrease when reproduced in a questioned impression.

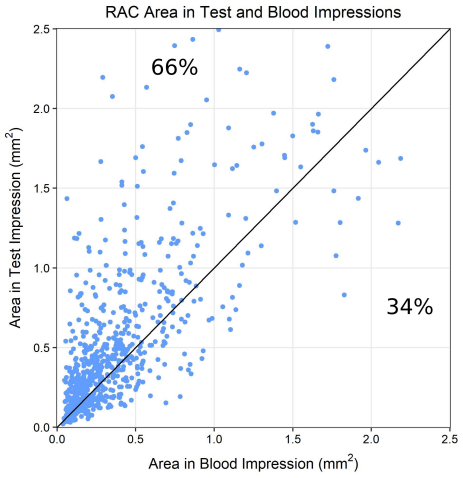


Figure 10: Relative area of 717 (94%) of the 759 RACs in test versus blood impressions (please see supplemental material section S3 and Fig. S4 for plot of all data including RAC areas greater than 2.5 mm^2). The diagonal line represents no change in RAC size between impression types.

The category of mated RACs was compared to assess changes in shape. A total of 576 $RACs_{(B)}$ (76%) maintained the same category as their mated $RAC_{(TSB)}$, while 183 $RACs_{(B)}$ were classified into a different category due to changes in RAC shape when reproduced in blood on tiles. Table 2 reports the confusion matrix comparing RAC categories for blood and test impressions.

Table 2: Confusion matrix showing the original categories of the $RACs_{(TSB)}$ in the test impressions and the categories of the mated $RACs_{(B)}$ in the blood impressions.

		Blood Impression		
		Linear	**Compact**	**Variable**
Test Impression	**Linear**	164	5	28
	Compact	3	136	26
	Variable	61	60	276

3.4. Quality Control

3.4.1. Repeatability in Marking

The number of RACs identified in each of the duplicate markings was counted for the 17 impressions comprising the quality control dataset (please see supplemental material section S4 and Table S1 for detailed results). While there was variability in RAC detection, RAC identification was not under investigation, and instead, differences in RAC shape between repeated markings was the focus. Each pair of duplicate impressions was compared to determine the number of RACs that were identified in both markings, and a total of 46 RAC pairs were found across all impressions. Of the 92 RACs in these pairs, 14 were unbiased and 78 were biased, resulting in 43 pairs having at least one RAC with a biased shape. The versions of the RAC from each marking were compared using percent area overlap, and the minimum value was 22.8%, the maximum was 90.2%, and the average $\pm$ one standard deviation was 71.3% $\pm$ 15.1%. Values less than 100% are attributed to variability in marking on the part of the researcher, as the paired impressions were identical copies of each other, so medium and substrate influences were not a factor. This is a weakness of this study since the overall probability of indistinguishability of two non-mated RACs is a function of RAC size and shape, which varies on average as much as 30% due to marking alone.

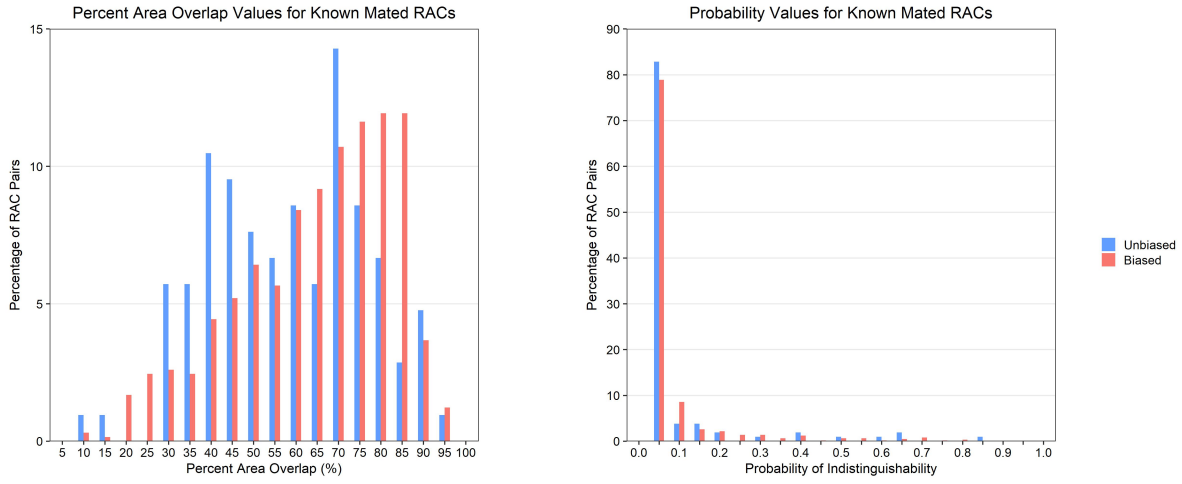
The effect of this %A difference on probability with respect to non-mated RAC pairs was further investigated. While a change in percent area overlap of 30% could have a drastic effect on probability for high percent area overlap values, approximately 90% of the non-mated RAC pairs have a %A value less than 50%. Table 3 reports the change in $p(\mathrm{I}|\%\mathrm{A})$ for each RAC category associated with five different 30% percent area overlap ranges. This demonstrates that within the relevant range of percent area overlap values for this study, the uncertainty in probability associated with researcher marking variability is less than 0.08.

Table 3: Change in $p(\text{I}|\%\text{A})$ for each RAC category model associated with different percent area overlap ranges of 30%.

	$p(\text{I}\|\%\text{A})$ Difference		
%A Range	**Linear**	**Compact**	**Variable**
0% – 30%	0.0133	0.0762	0.00418
5% – 35%	0.0185	0.0654	0.00428
10% – 40%	0.0248	0.0530	0.00358
15% – 45%	0.0336	0.0355	0.00240
20% – 50%	0.0481	0.0206	0.00080

3.4.2. Bias in Marking

The known mated RACs were separated into two groups based on whether or not the researcher could exhibit bias at the time of marking. Fig. 11(a) shows the distribution of the percent area overlap values of mated RAC pairs, where the x-axis indicates the inclusive upper limit of each bin. From this plot, it is clear that biased marking resulted in higher percent area overlap values. This result was confirmed with a chi-square test of independence [15], which generated a p-value of approximately 0.011 indicating a significant difference at a significance level of $\alpha = 0.05$. Fig. 11(b) shows the distribution of the probability of indistinguishability values of the same pairs. When the unbiased and biased distributions were compared using a chi-square test of independence [15], a p-value of approximately 0.44 was obtained, therefore failing to detect a statistically significant difference. This concludes that, while percent area overlap can be affected by biased marking, the downstream effect on probability as defined in this study is minor, and the potential influence of bias does not carry forward to impact RAC-RMF.



(a) (b)

Figure 11: Distribution of percent area overlap (a) and probability of indistinguishability (b) values for known mated RAC pairs, divided into two groups based on whether the shape of the RAC in the simulated crime scene impression could have been biased by the shape of the RAC in the test impression.

3.5. Known Non-Mated RAC Pairs

$\text{RACs}_{(\text{B})}$ from the blood impressions and their known mated $\text{RACs}_{(\text{TSB})}$ in the test impression subset were compared to known non-mated $\text{RACs}_{(\text{T})}$ with positional similarity on 1,299 unrelated outsoles. The results of non-mated comparisons for the blood impressions (Fig. 5, red arrows) are discussed in detail, while the results of non-mated comparisons for the test impression subset (Fig. 5, purple arrows) and the full database of test impressions [11] (Fig. 5, green arrows) are provided as a point of reference.

A total of 77,566 non-mated RAC comparisons were performed for both the blood impression dataset and the test impression subset. Compared to the 3.9 million comparisons performed when analyzing the full database of high-quality test impressions [11, 12], this represents a more than 50× decrease. For both the test impression subset and the blood impressions, 1,602 visual comparisons were performed, resulting in $p(\text{I})$ values of 0.0 or 1.0. For the test impression subset, 49 RAC pairs were deemed indistinguishable by visual comparison ($t = 1.0$, or *certain* indistinguishability). Note that this includes 22 pairs previously deemed indistinguishable [12] based on the variation expected in replicate test impressions, and 27 new pairs which were added in order to account for the increased variation expected when comparing a crime scene impression with a higher-quality test impression (please see supplemental section S5 for a brief discussion of uncertainty in these pairs, and section S6 and Figs. S5 and S6 for examples). For the blood impressions, 71 RAC pairs were deemed indistinguishable ($t = 1.0$, or *certain* indistinguishability).

A mathematical model based on percent area overlap [12] was used to predict the probability of indistinguishability of all RAC pairs not visually-compared. The upper 95% credible interval of this model provided $p(\text{I}|\%\text{A})$ values between 0.0 and 1.0 for each RAC pair, where a higher probability indicates a greater chance of two RACs being judged indistinguishable. Probability thresholds of $t \geq 0.75$ and $t \geq 0.5$ were chosen to define *plausible* and *possible* indistinguishable pairs, respectively. For the test impression subset, one additional *possible* pair was predicted by the model using a threshold of $t \geq 0.5$. For the blood impressions, one *plausible* pair and an additional 11 *possible* pairs were detected. The distribution of all probabilities is shown in Table 4 for the full database of high-quality test impressions [11], the test impression subset, and the blood impressions. The majority of non-mated RAC pairs have a probability of indistinguishability close to 0.0 for all datasets.

Table 4: Distribution of probability values resulting from all comparisons, including $RACs_{(T)}$ versus $RACs_{(T)}$ (test impressions) [11], $RACs_{(TSB)}$ versus $RACs_{(T)}$ (test impression subset), and $RACs_{(B)}$ versus $RACs_{(T)}$ (blood impressions).

	Test Impressions		**Test Impression Subset**		**Blood Impressions**	
Probability	**Count**	**%**	**Count**	**%**	**Count**	**%**
{0.0}	89,419	2.25	1,553	2.00	1,531	1.97
(0.0-0.25)	3,875,532	97.7	75,953	97.9	75,905	97.9
[0.25-0.5)	800	0.0202	10	0.0129	47	0.0606
[0.5-0.75)	133	0.00335	1	0.00129	11	0.0142
[0.75-1.0)	22	0.000554	0	0	1	0.00129
{1.0}	2,181	0.0550	49	0.0632	71	0.0915
Total Comparisons	3,968,087 (Fig. 5, green arrows)	—	77,566 (Fig. 5, purple arrows)	—	77,566 (Fig. 5, red arrows)	—

The longest axis of each RAC in an indistinguishable pair was measured to serve as a proxy for RAC size. For the comparison of $RACs_{(B)}$ in blood impressions to non-mated $RACs_{(T)}$ in test impressions, the longest axis of the RACs ($RACs_{(B)}$ + $RACs_{(T)}$) in the 83 *possible* indistinguishable pairs ($t \geq 0.5$) had a range of approximately 0.4 mm to 3.7 mm with a median value of 1.0 mm. Similarly, the RACs in the 50 pairs from the test impression subset ($RACs_{(TSB)}$ + $RACs_{(T)}$) had a range of approximately 0.4 mm to 4.0 mm with a median value of 1.2 mm. The non-mated $RACs_{(T)}$ in the 2,336 pairs from the full database of test impressions included the smallest and largest RACs observed in an indistinguishable pair, ranging from approximately 0.2 mm to 9.4 mm with a median value of 0.9 mm [11]. Fig. 12 plots the frequency (or percentage) of the longest axis measurements of RACs in indistinguishable pairs for all three sets of comparisons, where the x-axis indicates the inclusive upper limit of each bin. Selecting 2.0 mm as the longest axis measurement beyond which the majority of RACs (all RACs in the blood impression and test impression subset pairs and 98% of RACs in the test impression pairs) were linear features, the cumulative percentage of RACs with a length of at least 2.0 mm was 14.3% for the blood impressions, 22.0% for the test impression subset, and 17.8% for all test impressions.

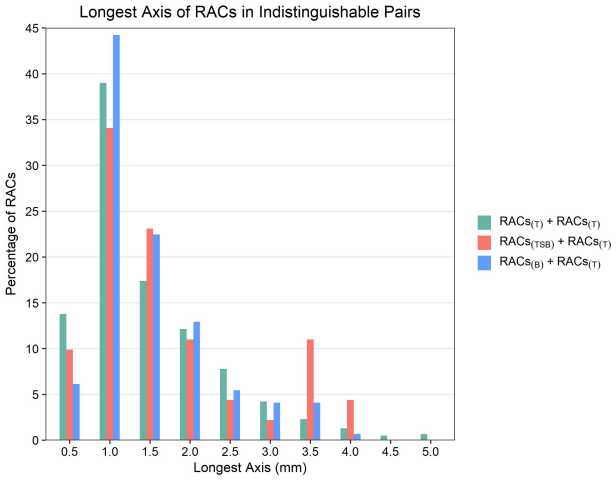


Figure 12: Distribution of longest axis measurements of RACs in *possible* indistinguishable pairs ($t \geq 0.5$) for the full database of test impressions ($\text{RACs}_{(\text{T})}$ + $\text{RACs}_{(\text{T})}$) [11], the test impression subset ($\text{RACs}_{(\text{TSB})}$ + $\text{RACs}_{(\text{T})}$), and the blood impressions ($\text{RACs}_{(\text{B})}$ + $\text{RACs}_{(\text{T})}$). Only RACs shorter than or equal to 5.0 mm are plotted, but this includes all RACs in the blood impression and test impression subset pairs, and 99% of the RACs in the pairs from the full database of test impressions.

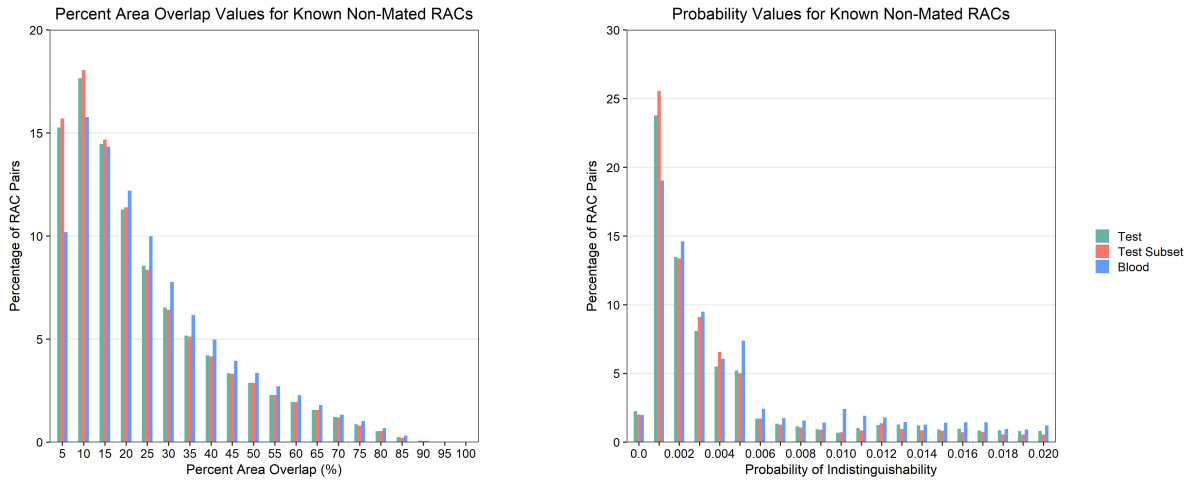
The percent area overlap values of the 71 *certain* indistinguishable non-mated RAC pairs shared between the blood impressions and non-mated test impressions ranged from 27.4% to 91.2%, with an average ± one standard deviation (SD) of 66.2% ± 17.4%. The percent area overlap of the single *plausible* indistinguishable pair was 85.3%. For the 11 additional pairs detected at $t \geq 0.5$, the minimum percent area overlap value was 74.5%, the maximum was 92.9%, and the average ± one standard deviation was 78.3% ± 5.20%. For comparison, the percent area overlap values from the test impressions [11] and the test impression subset are presented alongside the results from the blood impressions for all thresholds in Table 5. Overall, the percent area overlap values are reasonably similar for all impression categories within each threshold, with the test impression subset (center) exhibiting a slight increase in non-mated pairs with probabilities between 0.75 and 0.5, most likely explained by the drop out in pairs between 1.0 and 0.75.

Table 5: Percent area overlap values of indistinguishable pairs for all test impressions [11] (left), the test impression subset (center), and blood impressions (right) for each threshold.

				Percent Area Overlap (%)								
Probability	**RAC Pairs**			**Minimum**			**Maximum**			**Average ± 1 SD**		
$t = 1.0$	2,181	49	71	12.3	25.7	27.4	100	92.9	91.2	71.5 ± 16.0	70.1 ± 16.9	66.2 ± 17.4
$1.0 > t \geq 0.75$	22	0	1	83.1	—	85.3	97.9	—	85.3	88.6 ± 4.40	—	85.3 ± 0.00
$0.75 > t \geq 0.5$	133	1	11	74.4	92.2	74.5	93.0	92.2	92.9	78.5 ± 4.42	92.2 ± 0.00	78.3 ± 5.20

Fig. 13 shows the distribution of percent area overlap (13(a)) and the resulting probability of indistinguishability (13(b)) values for non-mated RAC pairs in the full database of high-quality test impressions [11], the test impression subset, and the blood impressions. Fig. 13(b) is limited to probability values in the range of [0.0-0.2], which captures approximately

75% of the data for each dataset. In both plots, the x-axis indicates the inclusive upper limit of each bin. These plots show overall similar trends between the datasets, with the most similar distribution observed between the test impressions and the test impression subset. While the $\text{RACs}_{(\text{B})}$ from the blood impressions also show a similar distribution, there is a noticeable shift towards higher percent area overlap and probability values. This indicates that $\text{RACs}_{(\text{B})}$ are more similar in geometry to non-mated $\text{RACs}_{(\text{T})}$ than when comparing non-mated test impressions to each other. In other words, "something" about the $\text{RACs}_{(\text{B})}$ in the blood impressions makes them more similar to non-mated $\text{RACs}_{(\text{T})}$ in unrelated test impressions than if comparing $\text{RACs}_{(\text{TSB})}$ to $\text{RACs}_{(\text{T})}$.



(a) (b)

Figure 13: Distribution of percent area overlap (a) and probability of indistinguishability (b) values for known non-mated RAC pairs from different impression datasets.

The percentage of indistinguishable pairs of each RAC category is shown in Table 6 for all thresholds. RAC categories also followed a similar trend for all three datasets. For visual comparisons ($t = 1.0$), linear RAC pairs were the most common, followed by compact, mixed, and variable. For indistinguishable pairs detected at lower thresholds by the mathematical model, linear and variable RAC pairs were still the most and least common, respectively, but mixed pairs were deemed indistinguishable more often than compact.

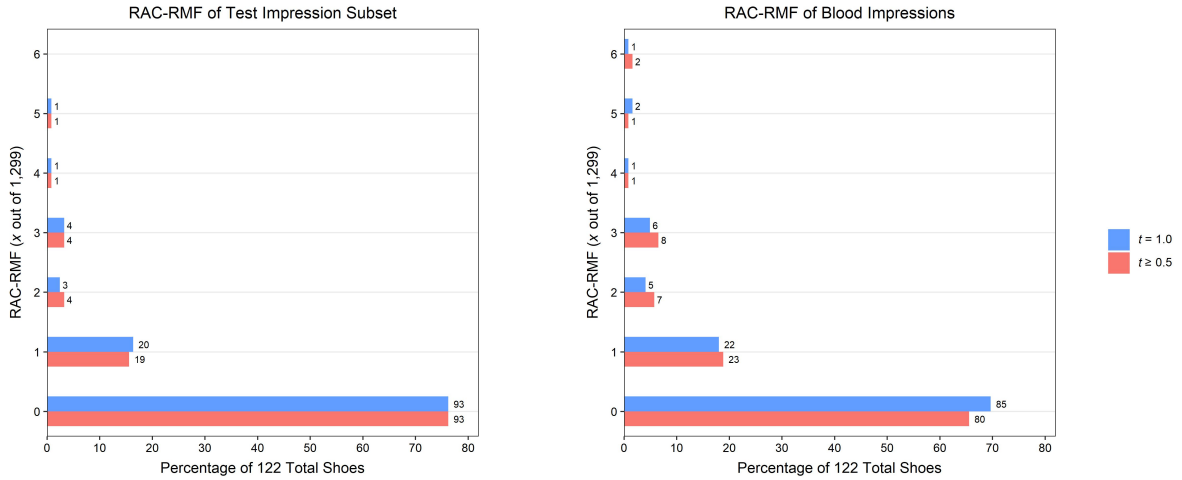
Table 6: Percentage of indistinguishable pairs of each RAC category for all test impressions [11] (left), the test impression subset (center), and blood impressions (right) for each threshold.

Probability	**RAC Pairs**			**Percentage of Total Pairs (%)**											
				Linear			**Compact**			**Variable**			**Mixed**		
$t = 1.0$	2,181	49	71	55.6	61.3	66.2	39.0	34.7	32.4	1.7	2.0	0	3.7	2.0	1.4
$1.0 > t \geq 0.75$	22	0	1	72.7	—	100	9.1	—	0	0	—	0	18.2	—	0
$0.75 > t \geq 0.5$	133	1	11	58.6	0	45.5	6.8	0	9.0	0	0	0	34.6	100	45.5

The spatial distribution of the indistinguishable pairs across the outsole was also investigated. For the blood impressions, the 71 visually-confirmed indistinguishable pairs were found in 47 different cells, with a maximum of four pairs in a single cell. The *plausible* indistinguishable pair detected at $t \geq 0.75$ was in the lateral heel. At $t \geq 0.5$, the 11 *possible* indistinguishable pairs were found in 11 different cells, none of which were the same as the location of the indistinguishable pairs with probabilities greater than 0.75. Spatial autocorrelation was not computed due to small sample size, but a trend was not visually apparent (please see supplemental material section S7 and Figs. S7–S9 for additional information).

3.6. Random Match Frequency of Randomly Acquired Characteristics

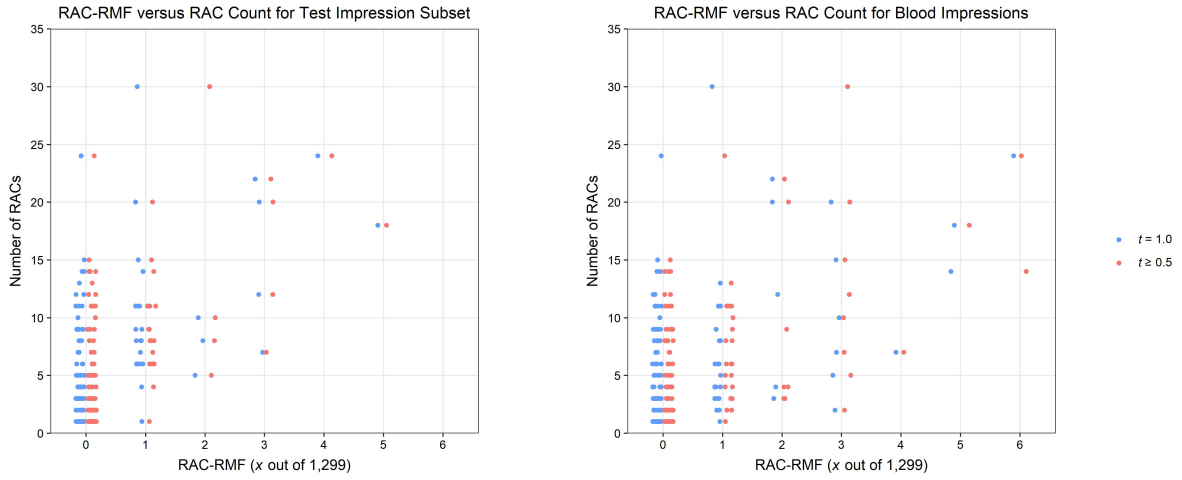
RAC-RMF$_{(m|n \geq 1)}$ was computed for each of the 122 shoes in the simulated crime scene impression dataset with at least one RAC that reproduced in a blood impression. Bar plots of the RAC-RMF$_{(m|n \geq 1)}$ values for these shoes based on the RACs in the blood impressions and the test impression subset are shown in Fig. 14. RAC-RMF$_{(m|n \geq 1)}$ is displayed as a value out of 1,299 on the y-axis, and the number of shoes corresponding to each RAC-RMF value is shown both as a percentage out of 122 on the x-axis and as a count to the right of each bar for $t = 1.0$ and $t \geq 0.5$. For the blood impressions at $t = 1.0$, 85 shoes (69.7%) did not share any indistinguishable RACs with an unrelated shoe in the database. The maximum RAC$_{(B)}$-RMF$_{(m|n \geq 1)}$ value observed was 6 out of 1,299 for one shoe (0.8%), meaning that one shoe shared at least one indistinguishable RAC pair with six unrelated outsoles. For $t \geq 0.5$, 80 shoes (65.6%) did not share any indistinguishable RACs with another shoe, while two shoes (1.6%) exhibited the maximum RAC$_{(B)}$-RMF$_{(m|n \geq 1)}$ value of 6 out of 1,299. When analyzing the mated RACs$_{(TSB)}$ from the test impression subset, 76.2% of shoes did not share an indistinguishable pair with any of the 1,299 unrelated shoes and the maximum RAC$_{(TSB)}$-RMF$_{(m|n \geq 1)}$ was 5 out of 1,299 for both thresholds.



(a) (b)

Figure 14: RAC-RMF$_{(m|n \geq 1)}$ as a value out of 1,299 for $t = 1.0$ and $t \geq 0.5$ for the test impression subset (a) and the blood impressions (b). As an example, 7 out of 122 (5.7%) blood impressions shared at least one indistinguishable pair with 2 out of 1,299 unrelated outsoles in the full database at $t \geq 0.5$.

Fig. 15 shows plots of the RAC-RMF$_{(m|n \geq 1)}$ values as a function of the number of RACs per shoe for the test impression subset (RACs$_{(\mathrm{TSB})}$) (15(a)) and the blood impressions (RACs$_{(\mathrm{B})}$) (15(b)) for $t = 1.0$ and $t \geq 0.5$. At most, there is a weak upward trend suggesting that shoes with fewer RACs tend to have lower random match frequencies than those with a greater number of RACs. This is likely explained by the fact that as the number of RACs in the impression increases, the chance of finding a similar RAC in the same position on another outsole also increases.



(a) (b)

Figure 15: RAC-RMF$_{(m|n \geq 1)}$ as a function of RAC count per shoe for the RACs$_{(\mathrm{TSB})}$ in the test impression subset (a) and RACs$_{(\mathrm{B})}$ in the blood impressions (b) for $t = 1.0$ and $t \geq 0.5$. The thresholds were dodged left and right by 0.4, and data points within each threshold were jittered horizontally $\pm$ 0.3 to improve visualization.

The occurrence of multiple indistinguishable RAC pairs shared between a pair of unrelated outsoles was also investigated. For the blood impressions and the subset of test impressions, no more than one indistinguishable RAC pair was ever shared between a pair of unrelated outsoles, meaning the maximum value of n in the expression RAC-RMF$_{(m|n)}$ was $n = 1$. This is in contrast to the full database of high-quality test impressions, where up to three RAC pairs ($n = 3$) for $t = 1.0$ and $t \geq 0.75$ and five RAC pairs ($n = 5$) for $t \geq 0.5$ were shared between a pair of unrelated outsoles [11].

3.6.1. Forensic Implications of Known Non-Mated RAC Pairs

In order to evaluate the reliability or quality of the RACs being compared, the variable of "size" was further explored. To estimate the forensic value associated with the RACs deemed indistinguishable in this study, a blood impression analyzed by approximately 70 subject matter experts in a recent footwear reliability study [17] was reviewed. When comparing this blood impression with its known mated test impression, approximately 50% of respondents reached the highest levels of source association afforded a footwear examiner when using the Scientific Working Group for Shoeprint and Tire Tread Evidence (SWGTREAD) 2013

conclusion standard (high degree of association or identification) [18]. Of the examiners not willing to make such a strong association, 27% associated class, while 16% erroneously formed the strongest disassociation possible (exclusion) [17]. Interestingly, those reporting an exclusion often based their opinion on an erroneous judgment of a size difference between the questioned and known mated test impression [17]. Conversely, those examiners reaching strong associations annotated eight major features, including three wear patterns and five RACs (please see supplemental section S8 and Fig. S10 for additional information). Further evaluation of the size of the five RACs based on their longest axis showed that the minimum was approximately 0.7 mm in length, the maximum was approximately 4.6 mm in length, and the average ± one standard deviation was 2.8 mm ± 1.5 mm. Using this study [17] as a point of reference, more than 90% of the RACs detected in the blood impressions and in the test impression subset in the present study were larger than 0.7 mm. Using the average length as an approximation for RAC size that examiners deem reliable in a comparison, 131 (17%) of the RACs detected in the blood impressions and 280 (37%) of the RACs in the test impression subset had lengths greater than or equal to 2.8 mm. Table 7 summarizes the number and percentage of shoes with non-zero RAC-RMFs when the criteria of RAC length is greater than or equal to 2.8 mm, versus the same metric without regard for RAC length. If only larger/more-reliable RACs were considered when assessing indistinguishable pairs, the blood impression dataset would be reduced to five indistinguishable pairs and the test impression subset to eight indistinguishable pairs, all of which were visually-confirmed ($t = 1.0$). In terms of RAC-RMF, 3 (3.4%) out of 89 shoes for the blood impression dataset and 7 (7.9%) out of 89 shoes for the test impression subset would have a RAC-RMF$_{(m|n \geq 1)}$ greater than or equal to 1 out of 1,299, where 89 is the number of shoes with at least one RAC longer than or equal to 2.8 mm in the test impression subset. When considering RACs of any size, the proportions of shoes with non-zero RAC-RMFs are reported out of a possible 122 shoes.

Table 7: Percentage of shoes with non-zero RAC-RMFs including only those RACs with a length greater than or equal to 2.8 mm at $t = 1.0$ versus RACs of any length/size at $t \geq 0.5$. Shoe count is reported as a fraction of the number of shoes with at least one RAC meeting the RAC length criteria (based on the test impression subset), followed by the percentage in parentheses. The percentage change was calculated with reference to the test impression subset using (B – TSB)/TSB × 100%, where a negative value indicates a decrease (*e.g.*, using the first two rows as an example, 7 versus 3 shoes is a 57% decrease (3 – 7)/7 × 100% = -57%).

Comparison	**Probability**	**RAC Length**	**Shoes with RAC-RMF ≥ 1 out of 1,299**	**Percentage Change**
RACs$_{(TSB)}$ versus RACs$_{(T)}$	$t = 1.0$	≥ 2.8 mm	7/89 (7.9%)	-57%
RACs$_{(B)}$ versus RACs$_{(T)}$	$t = 1.0$	≥ 2.8 mm	3/89 (3.4%)	
RACs$_{(TSB)}$ versus RACs$_{(T)}$	$t \geq 0.5$	Any	29/122 (23.8%)	45%
RACs$_{(B)}$ versus RACs$_{(T)}$	$t \geq 0.5$	Any	42/122 (34.4%)	

4. Discussion

The overall goal of this study was to provide a lower-quality complement to previous research which estimated random match frequency of RACs based on high-quality test impressions [11]. To accomplish this, a subset of 165 shoes from the West Virginia University footwear database was used to create a dataset of simulated crime scene impressions made in blood on different tile substrates. After enhancement with a leucocrystal violet solution, each impression was registered to a test impression of the corresponding outsole (which decreased the number of usable impressions to 162), and was marked by a single researcher to identify RACs. A modified sequential unmasking approach was implemented, which included an initial blind marking of the blood impression followed by a secondary marking while viewing the previously-marked test impression. This resulted in biased shapes for those RACs identified after viewing the test impression. Using a chi-square test of independence, biased marking was found to influence the percent area overlap value during comparisons, but not significantly affect the probability of indistinguishability as computed in this study. As a consequence, results were not divided and separately analyzed as a function of whether or not the RAC could have been biased during marking.

A total of 759 $\text{RACs}_{(\text{B})}$ were identified in blood impressions created from 162 shoes known to possess 18,887 $\text{RACs}_{(\text{TB})}$, meaning that over 95% of the RACs from the outsoles did not transfer. Previously, a study of simulated crime scene impressions created using shoe polish as the impression medium reported that approximately 85% of the RACs did not transfer from shoes to the resulting impressions [19]. While the impressions created using shoe polish were of lower quality than Handiprint test impressions, the medium was able to capture more detail than the blood used in this study, which is supported by the 10% difference in the percentage of RACs transferred from the outsoles. In combination, both datasets demonstrated that with impressions of lower quality than Handiprints, a small proportion of RACs are likely to transfer. However, those RACs that do transfer can play an important role in forming the highest levels of source association. For this dataset, 75% of blood impressions possessed *at least* one RAC, while the average number of RACs transferred was five.

In terms of location, as compared to the high-quality test impressions, the spatial distribution of RACs in blood impressions showed that RACs around the perimeter of the outsole almost never transferred. However, it must be noted that the test impressions were created using a static benchtop method [10] while the blood impressions were created using a dynamic walking method [1]. As a result, it is hypothesized that differences in RAC density around the outsole perimeter are a function of the way the impressions were generated, but this remains a variable for future study. In addition to location, the size of the RACs that transferred was also investigated, revealing that larger RACs from the test impressions were more likely to transfer and be detected in the blood impressions than smaller RACs. This is most likely explained by the fact that larger RACs are easier for a researcher to identify, and that smaller RACs are more likely to be filled in by excess blood and/or deemed below the signal-to-noise limit when considering the medium and substrate interferences present in this study. Perhaps more interesting is that two-thirds (66%) of the RACs in the blood

impressions were physically smaller than their corresponding known mate in the test impressions (Fig. 10). This could be explained by edge erosion due to the use of a liquid medium to create the impressions and/or a byproduct of the enhancement technique, but should be further investigated to determine if this finding persists when based on impressions deposited in other media.

The results outlined in [11] focused only on comparisons of $\text{RACs}_{(\text{T})}$ versus $\text{RACs}_{(\text{T})}$. For the purposes of the present study, $\text{RACs}_{(\text{B})}$ were compared to $\text{RACs}_{(\text{T})}$ with positional similarity. While the shoes in both datasets were compared to 1,299 unrelated outsoles, the 3,968,087 observations in [11] represent 50× more comparisons than the 77,566 RAC pairs available for analysis in this study, resulting in many more chances to find at least one shared indistinguishable RAC pair between unrelated outsoles. This difference in magnitude made it difficult to draw conclusions between the two datasets. To combat this issue and provide a more comparable point of reference, the $\text{RACs}_{(\text{TSB})}$ from the test impressions that reproduced in blood impressions were used to form a test impression subset, resulting in another 77,566 comparisons, and therefore opinions based on a one-to-one correspondence in data of equal sample size.

When considering the number of *possible* ($t \geq 0.5$) indistinguishable pairs found out of the total number of comparisons conducted for each dataset, there is an approximately 1 in 1,700 frequency of encountering an *indistinguishable pair* for the full database of test impressions and a 1 in 1,550 frequency for the test impression subset. In contrast, there is an approximately 1 in 900 frequency of encountering an *indistinguishable pair* for the blood impression dataset. This represents a nearly two-thirds increase in the number of *possible* indistinguishable pairs ($(83 - 50)/50 \times 100\% = 66\%$) for the blood impressions analyzed in this study, which is based on the observation that blood impressions compared to unrelated test impressions resulted in higher percent area overlap and probability of indistinguishability values than either test impression dataset (Fig. 13). Although there are likely several confounding factors that could be associated with this observation, one possible explanation is based on the fact that $\text{RACs}_{(\text{B})}$ from blood impressions were overall smaller than the corresponding $\text{RACs}_{(\text{TSB})}$ from test impressions. While not always the case, it is reasonable to assume that "smaller" is associated with decreased complexity (especially when comparing binary digitized shapes), and therefore increased percent area overlap values.

$\text{RAC-RMF}_{(m|n \geq 1)}$ was computed for each of the 122 shoes in the simulated crime scene impression dataset that had at least one RAC in its mated blood impression. This process was performed using $\text{RACs}_{(\text{B})}$ in the blood impressions and $\text{RACs}_{(\text{TSB})}$ in the corresponding test impressions to determine the number of unrelated shoes out of 1,299 that shared *at least one* indistinguishable RAC pair with a held-out shoe. Comparison of the results of the test impression subset with the full database of test impressions highlights the effect of RAC count on RAC-RMF. With 50× fewer comparisons (3.9M versus 77.5K), the maximum $\text{RAC-RMF}_{(m|n \geq 1)}$ decreased from 49 out of 1,299 [11] to 5 out of 1,299. In addition, the proportion of shoes with a non-zero RAC-RMF was reduced from nearly 70% in the full database of test impressions [11] to 24% in the test impression subset.

By controlling for the difference in RAC count and number of comparisons, the results of the test impression subset and blood impression dataset are more easily compared, and

the influence of impression type on similarity in non-mated RAC pairs can be examined. In the absence of any size/length threshold, the larger number of indistinguishable pairs for the blood impression dataset compared to the test impression subset directly contributed to higher $\text{RAC}_{(\text{B})}\text{-RMF}_{(m|n \geq 1)}$ values overall. This was observed most notably by a 45% increase (at a threshold of $t \geq 0.5$) in the number of shoes with a RAC deemed indistinguishable to an unrelated outsole (42/122 for blood versus 29/122 for the test impression subset). This demonstrates that $\text{RACs}_{(\text{B})}$ are more similar to $\text{RACs}_{(\text{T})}$ than $\text{RACs}_{(\text{TSB})}$ are to $\text{RACs}_{(\text{T})}$, leading to an overall greater chance of finding indistinguishable pairs between blood impressions and non-mated outsoles. This observation reverses when a RAC length greater than or equal to 2.8 mm is required. At this threshold, there was a 57% decrease in shoes with non-zero RAC-RMFs when impressions are deposited in blood on tile versus laboratory-prepared test impressions (3/89 for blood versus 7/89 for the test impression subset).

4.1. Limitations

As previously discussed in [11], there were several notable limitations of that study which were not addressed and therefore persist in the present study, including mixed makes and models of shoes, the use of an outsole normalization procedure, and the use of a binning process to denote positional similarity. Another potential limitation was introduced in this study with regard to the use of percent area overlap as a similarity metric. Previous work evaluating different similarity metrics for the comparison of RACs from simulated crime scene impressions made in shoe polish versus test impressions concluded that Hausdorff distance had better performance than percent area overlap [12]. However, the use of Hausdorff distance for the purposes of the present study would prevent direct comparisons to the results previously reported in [11], so percent area overlap was used despite the acknowledged limitation.

In addition to the above limitations, there are four more variables that were not addressed in this study, and therefore the subsequent effect of these on the interpretation of the results is unknown. First, the use of a modified sequential unmasking procedure during marking prevented the inclusion of pseudo-accidentals in this dataset (where "pseudo" can mean erroneously identifying RACs that are not actually present, and/or erroneously inferring a shape or geometry for one or more RACs). As a result, the impact of pseudo-accidentals on RAC-RMF estimates cannot be quantified. Second, it must be stated that the focus of this study was to evaluate non-mated RAC pairs from impressions of different quality in order to inform RAC-RMF estimates, but no attempt was made to translate this into an estimate of a likelihood ratio (LR). This decision is based on two primary obstacles. The first barrier is that the mathematical model [12] used for predicting probability of indistinguishability was trained using non-mated pairs, and therefore analysis of known mates would require a different model trained by the results of visual comparisons of mated pairs. The second concern is that only one simulated crime scene replicate was created for each shoe, which leads to a small sample size of mated pairs to inform the numerator in an LR. Previous research has shown that a RAC's shape can vary between replicate test impressions [20], so multiple replicates are certainly required to characterize this variation in simulated crime

scene impressions. As a result, this study focused only on the denominator in an LR, and exploration of the numerator is an avenue for future consideration. Third, RAC-RMF estimates were based only on the occurrence of indistinguishable RAC pairs (n), with no regard for the number of distinguishable pairs (m) shared between unrelated outsoles. This is an important distinction, as the strength of an association between questioned and known impressions during a forensic comparison would require an impression-wide assessment of RMF, which must consider RACs that are dissimilar or disagree, in addition to those in agreement.

Fourth, this study did not thoroughly investigate the *reliability* or *quality* of the RACs being compared. At most, the attribute of "size" was considered. Based on the RACs annotated by examiners when comparing a blood impression on tile to its mated high-quality test impression in a recent black box study [17], the average length of the RACs used to reach the highest levels of source association was 2.8 mm. For the impressions generated for this study, 17% (131/759) of RACs in blood impressions were above this threshold, demonstrating the potential utility of these RACs during a forensic comparison. Based on a conservative estimate for RAC size as a substitute for reliability, at least three shoes in this study presented a non-zero $\text{RAC}_{(\text{B})}\text{-RMF}_{(m|n \geq 1)}$ for features greater than or equal to 2.8 mm in length at $t = 1.0$. Using a more permissive criteria, this estimate could be as high as 42 shoes when including smaller RACs at $t \geq 0.5$.

4.2. Forensic Implications

In summary, this investigation found that without regard for RAC length, examiners can anticipate that $\text{RAC}_{(\text{B})}$-RMFs are greater than $\text{RAC}_{(\text{TSB})}$-RMFs due to the increased degree of geometric similarity between $\text{RACs}_{(\text{B})}$ and non-mated $\text{RACs}_{(\text{T})}$. As previously noted, this observation is believed to be a function of the decreased complexity associated with digitized geometric shapes as they decrease in overall size. When RAC length is used as a proxy for reliability and only larger RACs are considered, this trend reverses. Although the RMF of a questioned impression cannot be deduced solely from RAC-RMF estimates, this research shows that RACs in simulated questioned impressions of the type and quality expected in casework can and do co-occur in position and geometry with RACs in non-mated test impressions at a rate of 3.4% (3 out of 89 shoes with one or more RACs greater than or equal to 2.8 mm in length) with a relative random match frequency of at least 0.0008 (1 out of 1,299). Since theoretical models have traditionally been the basis for estimating RAC-RMFs in footwear comparisons, this research allows for calibration based on empirical data, thereby advancing forensic foundational knowledge. In order to determine if this estimate is robust, and if the preceding hypotheses remain reasonable, this type of study should be independently repeated using different datasets and similarity metrics, as well as different media and substrates. Part II of this investigation addresses the latter variables by performing an analogous investigation using dust impressions deposited on paper and tile, with those on tile lifted using Mylar film or gelatin.

Acknowledgments

The database used in this investigation was originally supported by Award No. 2013-DN-BX-K043, awarded by the National Institute of Justice (NIJ), Office of Justice Program, U.S. Department of Justice. The simulated crime scene dataset was supported by the Center for Statistics and Applications in Forensic Evidence (CSAFE), through Cooperative Agreement No. 70NANB20H019 between NIST and Iowa State University, which includes activities carried out at West Virginia University. Images can be found in the CSAFE repository and are cited to Smale, A., Speir, J., West Virginia University 2022 High Quality and Simulated Crime Scene Dataset, Release #1; Release Date March 2022. In addition to the NIJ and CSAFE, the authors would like to thank Claire Dolton for registering the 165 blood impressions analyzed in this investigation. Lastly, the opinions, findings, conclusions, and recommendations expressed in this manuscript are those of the authors and do not necessarily reflect those of the Department of Justice, the Center for Statistics and Applications in Forensic Evidence, and/or the National Institute of Standards and Technology.

References

[1] W. J. Bodziak, Footwear Impression Evidence: Detection, Recovery, and Examination, 2nd Edition, CRC Press, Boca Raton, FL, 2000.

[2] M. J. Cassidy, Footwear Identification, Public Relations Branch of the Royal Canadian Mounted Police, Ottawa, Ontario, 1980.

[3] T. W. Adair, J. LeMay, A. McDonald, R. Shaw, R. Tewes, The Mount Bierstadt study: An experiment in unique damage formation in footwear, Journal of Forensic Identification 57 (2) (2007) 199–205.

[4] C. Hamburg, R. Banks, Evaluation of the random nature of acquired marks on footwear outsoles, Impression and Pattern Evidence Symposium; Clearwater, FL (2010).

[5] H. D. Wilson, Comparison of the individual characteristics in the outsoles of thirty-nine pairs of Adidas Supernova Classic shoes, Journal of Forensic Identification 62 (3) (2012) 194–203.

[6] M. Marvin, A look at close non-match footwear examinations, International Association for Identification (IAI) Centennial Conference; Cincinnati, OH (2015).

[7] A. S. Fawcett, The role of the footmark examiner, Journal of the Forensic Science Society 10 (4) (1970) 227–244. doi:https://doi.org/10.1016/S0015-7368(70)70613-0.

[8] R. S. Stone, Footwear examinations: Mathematical probabilities of theoretical individual characteristics, Journal of Forensic Identification 56 (4) (2006) 577–599.

[9] Y. Yekutieli, Y. Shor, S. Wiesner, T. Tsach, Expert assisting computerized system for evaluating the degree of certainty in 2D shoeprints, Technical Report. National Institute of Justice, U.S. Department of Justice, Office of Justice Programs (2012).

[10] J. A. Speir, N. Richetelli, M. Fagert, M. Hite, W. J. Bodziak, Quantifying randomly acquired characteristics on outsoles in terms of shape and position, Forensic Science International 266 (2016) 399–411. doi:https://doi.org/10.1016/j.forsciint.2016.06.012.

[11] A. N. Smale, J. A. Speir, Estimate of the random match frequency of acquired characteristics in a forensic footwear database, Science & Justice 63 (3) (2023) 427–437. doi:https://doi.org/10.1016/j.scijus.2023.04.007.

[12] N. Richetelli, W. J. Bodziak, J. A. Speir, Empirically observed and predicted estimates of chance association: Estimating the chance association of randomly acquired characteristics in footwear comparisons, Forensic Science International 302 (2019) 1–14. doi:https://doi.org/10.1016/j.forsciint.2019.05.049.

[13] E. T. Lin, T. DeBat, J. A. Speir, A simulated crime scene footwear impression database for teaching and research purposes, Journal of Forensic Sciences 67 (2022) 726–734. doi:https://doi.org/10.1111/1556-4029.14933.

[14] I. E. Dror, W. C. Thompson, C. A. Meissner, I. Kornfield, D. Krane, M. Saks, M. Risinger, Letter to the editor- Context management toolbox: A linear sequential unmasking (LSU) approach for minimizing cognitive bias in forensic decision making, Journal of Forensic Sciences 60 (4) (2015) 1111–1112. doi:https://doi.org/10.1111/1556-4029.12805.

[15] M. L. McHugh, The Chi-square test of independence, Biochemia medica 23 (2) (2013) 143–149. doi:https://doi.org/10.11613/BM.2013.018.

[16] F. Garwood, Fiducial limits for the Poisson distribution, Biometrika 28 (3/4) (1936) 437–442. doi:https://doi.org/10.2307/2333958.

[17] N. Richetelli, L. Hammer, J. A. Speir, Forensic footwear reliability: Part II—Range of conclusions, accuracy, and consensus, Journal of Forensic Sciences 65 (6) (2020) 1871–1882. doi:https://doi.org/10.1111/1556-4029.14551.

[18] SWGTREAD, Range of Conclusions Standard for Footwear and Tire Impression Examinations, https://www.nist.gov/system/files/documents/2016/10/26/swgtread_10_range_of_conclusions_standard_for_footwear_and_tire_impression_examinations_201303.pdf, 2013 (accessed June 2023).

[19] N. Richetelli, M. Nobel, W. J. Bodziak, J. A. Speir, Quantitative assessment of similarity between randomly acquired characteristics on high quality exemplars and crime scene impressions via analysis of feature size and shape, Forensic Science International 270 (2017) 211–222. doi:https://doi.org/10.1016/j.forsciint.2016.10.008.

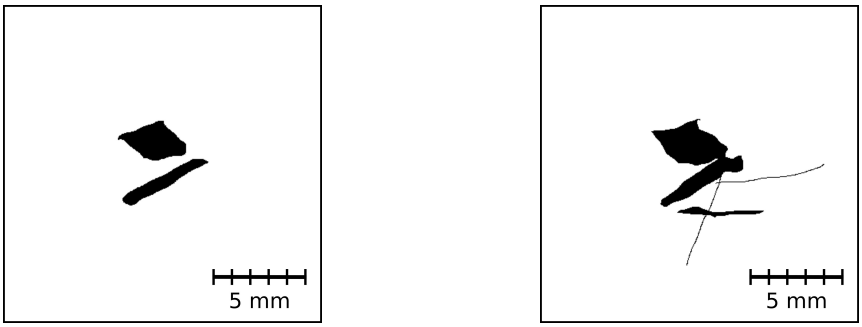
[20] Y. Shor, S. Wiesner, T. Tsach, R. Gurel, Y. Yekutieli, Inherent variation in multiple shoe-sole test impressions, Forensic Science International 285 (2018) 189–203. doi:https://doi.org/10.1016/j.forsciint.2017.10.030.

Estimate of the Random Match Frequency of Acquired Characteristics in Footwear: Part I — Simulated Crime Scene Impressions in Blood

Supplemental Material

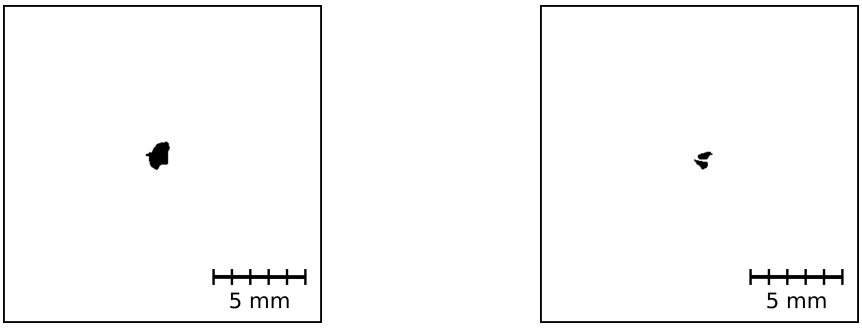
S1. Removed RACs from the Blood Impressions

A total of 782 RACs were originally marked and confirmed in the blood impressions. However, after further inspection of the RACs from the corresponding test impressions, 23 of these RACs were removed (out of convenience) for one of two reasons. The first scenario, shown in Fig. S1, occurred when multiple RACs identified in the blood impression (S1(a)) originated from conjoined RACs treated as a single connected component in the test impression (S1(b)). The second scenario, shown in Fig. S2, occurred when multiple RACs from the test impression (S2(b)) merged into a single RAC in the blood impression (S2(a)).



(a) (b)

Figure S1: Two separate RACs in the blood impression (a) and multiple conjoined RACs treated as a single connected component in the test impression (b). The authors acknowledge that (b) includes five different RACs, but this would require manual separation following the connected components extraction.

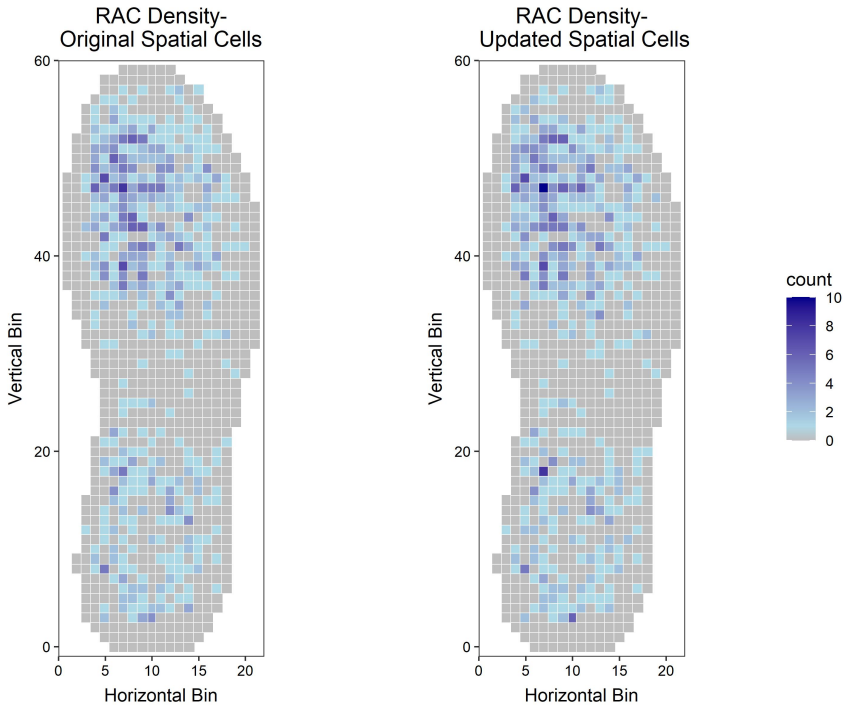


(a) (b)

Figure S2: One RAC in the blood impression (a) that resulted from two RACs in the test impression merging together (b).

S2. Spatial Distribution of RACs in the Blood Impressions

Each of the 759 marked $\text{RACs}_{(\text{B})}$ from the blood impressions was extracted and mapped to the normalized outsole based on coordinate information, as outlined by Speir *et al.* (2016). These $\text{RACs}_{(\text{B})}$ were distributed in 406 out of 987 spatial cells with maximum of 8 RACs per cell, as shown in Fig. S3(a). Upon comparison with the mated $\text{RAC}_{(\text{TSB})}$ in the test impression, 105 (14%) of the $\text{RACs}_{(\text{B})}$ in the blood impressions were shifted into a neighboring spatial cell. This occurred for two possible reasons; misalignment in co-registration of the blood impression to its mated test impression, and/or a change in the RAC's (x, y)-centroid (which determines the appropriate spatial cell) as a result of changes in size and/or shape of the RAC in the blood impression. In order to ensure a consistent set of comparisons between the $\text{RACs}_{(\text{B})}$ in the blood impressions and those in the test impression subset, the 105 misassigned $\text{RACs}_{(\text{B})}$ from the blood impressions were remapped, treating the test impression centroid as ground truth. After updating, the $\text{RACs}_{(\text{B})}$ were distributed in 394 out of 987 spatial cells with a maximum of 10 RACs per cell, as shown in Fig. S3(b). These spatial cells were used to determine the necessary RAC comparisons moving forward.



(a) (b)

Figure S3: Distribution of $\text{RACs}_{(\text{B})}$ from the blood impressions in the original spatial cells determined by the RAC's position as marked (a) and in the updated spatial cells using the mated $\text{RAC}_{(\text{TSB})}$ from the test impression as ground truth (b).

S3. Known Mate RAC Areas

The area of each RAC in the blood impressions and their mates in the corresponding test impressions was computed, and the results are shown in Fig. S4. Data points above the diagonal line (66%) represent RACs in the blood impressions that were smaller than their test impression mate, while those below the line (34%) represent RACs that were larger in the blood impression. Only three RACs (<0.5%) were the same size in both impressions. All 759 RACs are shown in this plot.

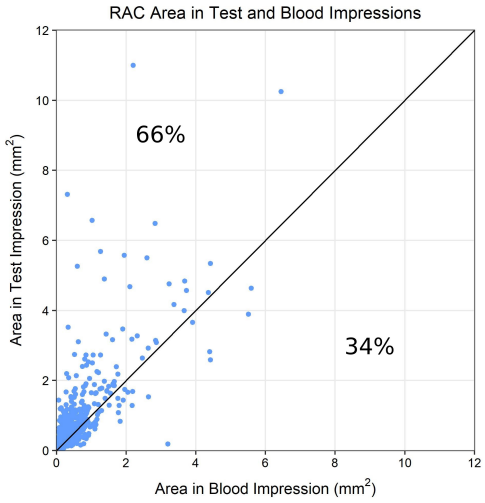


Figure S4: Relative area of known mated RACs in test and blood impressions. The diagonal line represents no change in RAC size between impression types.

S4. Quality Control

RAC count comparisons for the 17 impressions that were marked twice for quality control purposes are shown in Table S1. The researcher was blinded as to which images were assigned to the quality control testing program. However, during post-processing the duplicate impressions were denoted as "Marking A" and "Marking B" where "A" reflects the first time an impression was processed and "B" represents the second. Using the second row of the table for shoe 118R as an example, 20 RACs (19 of which were biased) were identified during the first mark-up (A) and 24 RACs (20 of which were biased) were identified in the second attempt (B), for a count difference of 4 RACs and a percentage difference out of the maximum number found of 83.3% (or (20/24) × 100%). When comparing the RACs identified during each marking, 15 RACs were consistent in both markings, which is 68.2% of the total number of RACs found (or (15+15)/(20+24) × 100%). In all but one case (shoe 474L), the number of RACs identified in the second marking was equal to or greater than the number identified in the first marking, which suggests consistent performance or possible improvement over time, although statistical testing was not conducted since RAC identification was not a variable under investigation in this study.

Table S1: RAC counts from duplicate markings of the 17 impressions used as part of the quality control program. For RAC and bias counts, the first number corresponds to Marking A and the second corresponds to Marking B.

Shoe ID	RAC Count (A/B)	Bias Count (A/B)	Count Difference	RACs in Both	Percentage of Max. (%)	Percentage in Both (%)
024L	0/0	0/0	0	0	—	—
118R	20/24	19/20	4	15	83.3	68.2
270L	0/2	0/1	2	0	0	0
276L	1/2	1/2	1	1	50.0	66.7
288L	3/3	3/3	0	2	100	66.7
296R	4/4	4/4	0	3	100	75.0
333R	0/0	0/0	0	0	—	—
474L	8/2	8/2	6	2	25.0	40.0
476R	5/9	3/8	4	3	55.6	42.9
491R	7/7	7/7	0	6	100	85.7
502R	8/13	7/11	5	8	61.5	76.2
513L	1/2	1/2	1	1	50.0	66.7
525R	5/5	3/2	0	5	100	100
566L	0/0	0/0	0	0	—	—
622R	0/0	0/0	0	0	—	—
631R	1/2	1/2	1	0	50.0	0
666R	0/0	0/0	0	0	—	—

S5. Uncertainty in Visual Comparisons

There were 1,602 visual comparisons performed between $\text{RACs}_{(\text{TSB})}$ and $\text{RACs}_{(\text{T})}$. These comparisons were performed a minimum of three times by two researchers. During these repeated trials, a total of 48 RAC pairs were consistently deemed indistinguishable. One additional pair was sometimes classified as indistinguishable and sometimes classified as exhibiting minor differences, and therefore was deemed uncertain. The same process was repeated when comparing $\text{RACs}_{(\text{B})}$ and $\text{RACs}_{(\text{T})}$. For these trials, 67 pairs were consistently deemed indistinguishable and four pairs were uncertain. In order to report a worst-case scenario, these uncertain pairs were counted as indistinguishable for all analyses (for a total of 49 and 71 pairs, respectively).

S6. Indistinguishable Pairs

Examples of RAC pairs deemed indistinguishable by visual comparison are shown in Figs. S5 and S6. A selection of the 22 RAC pairs previously deemed indistinguishably by Richetelli *et al.* (2019) using "Test Criteria" are shown in Fig. S5, where both RACs are from unrelated test impressions. A selection of additional indistinguishable pairs using the more lenient "CS Criteria" are shown in Fig. S6, where the black RAC is from a blood impression and the gray RAC is from an unrelated test impression.

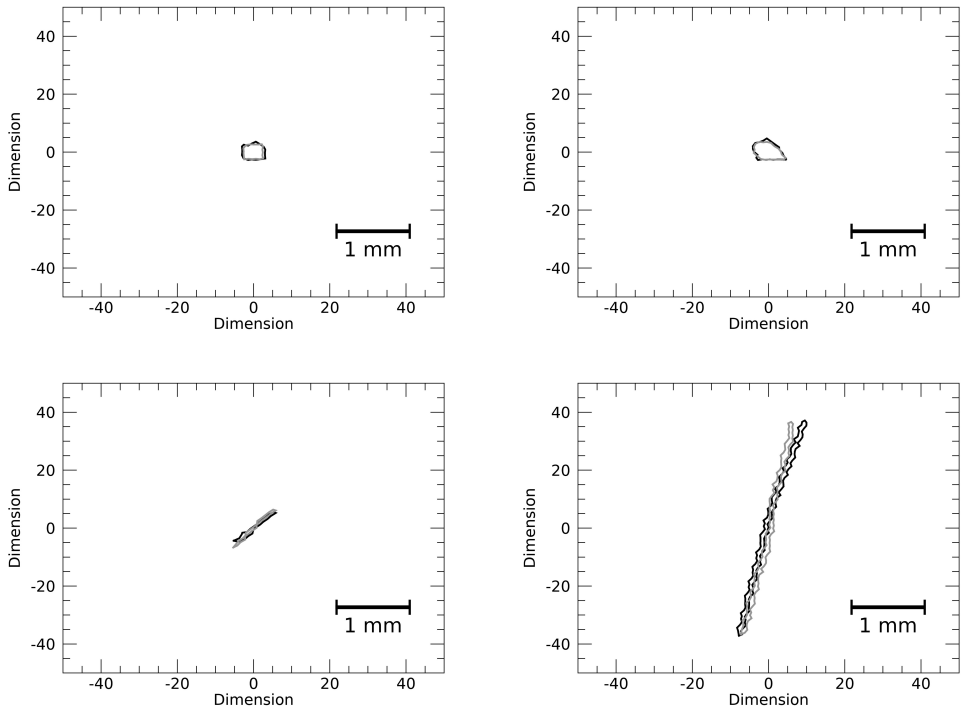


Figure S5: A sampling of non-mated $\text{RACs}_{(\text{T})}$ deemed indistinguishable using Test Criteria.

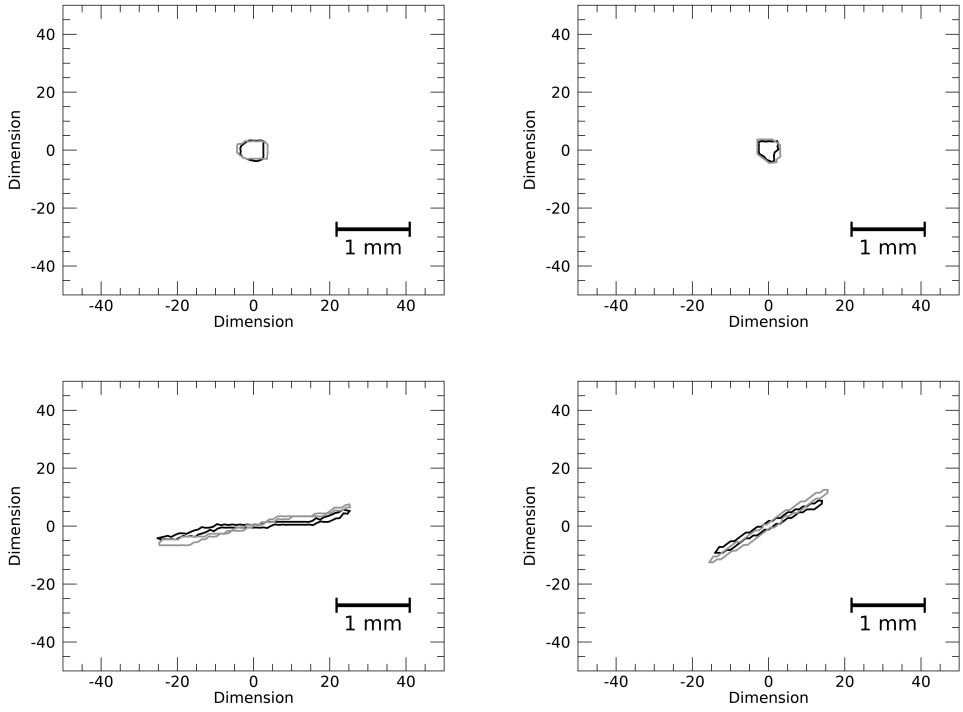
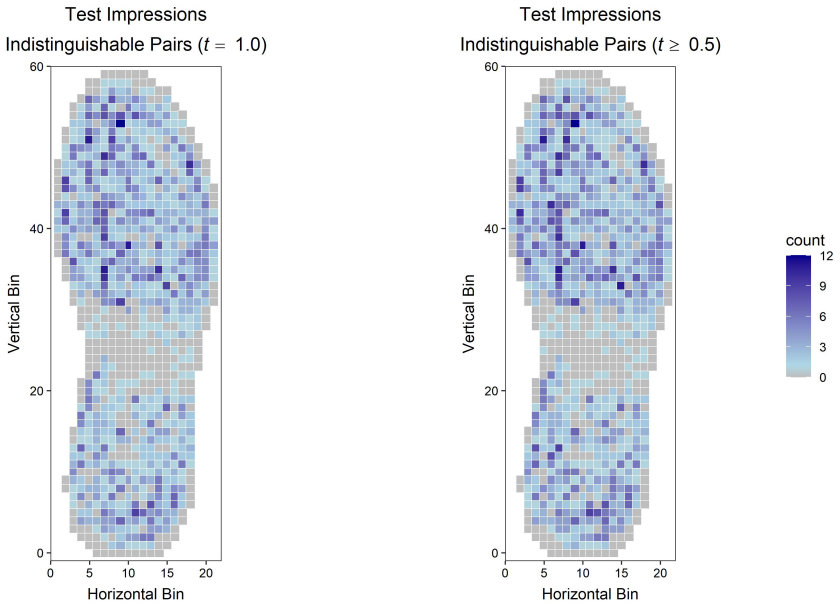


Figure S6: A sampling of $\text{RACs}_{(\text{B})}$ (black) and $\text{RACs}_{(\text{T})}$ (gray) deemed indistinguishable using CS Criteria.

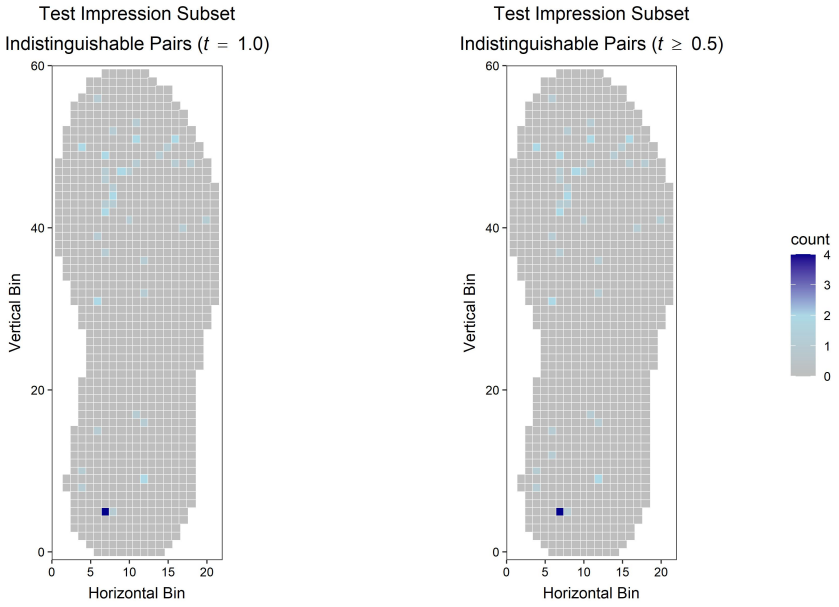
S7. Spatial Distribution of Indistinguishable Pairs

The spatial distribution of the indistinguishable RAC pairs is shown for the full database of test impressions (Fig. S7) as discussed by Smale and Speir (2023), the subset of RACs from the test impressions with a mate in the corresponding blood impression (Fig. S8), and the blood impressions (Fig. S9) for $t = 1.0$ and $t \geq 0.5$. Due to the similarity between the results for $t = 1.0$ and $t \geq 0.75$ for all three datasets, the heatmaps for $t \geq 0.75$ are not included. For the full database of test impressions, there were 2,181 indistinguishable pairs at a threshold of $t = 1.0$ and 2,336 pairs at $t \geq 0.5$. For the test impression subset, there were 49 indistinguishable pairs at a threshold of $t = 1.0$ and 50 pairs at $t \geq 0.5$. Lastly, for the blood impressions, there were 71 indistinguishable pairs at a threshold of $t = 1.0$ and 83 pairs at $t \geq 0.5$.



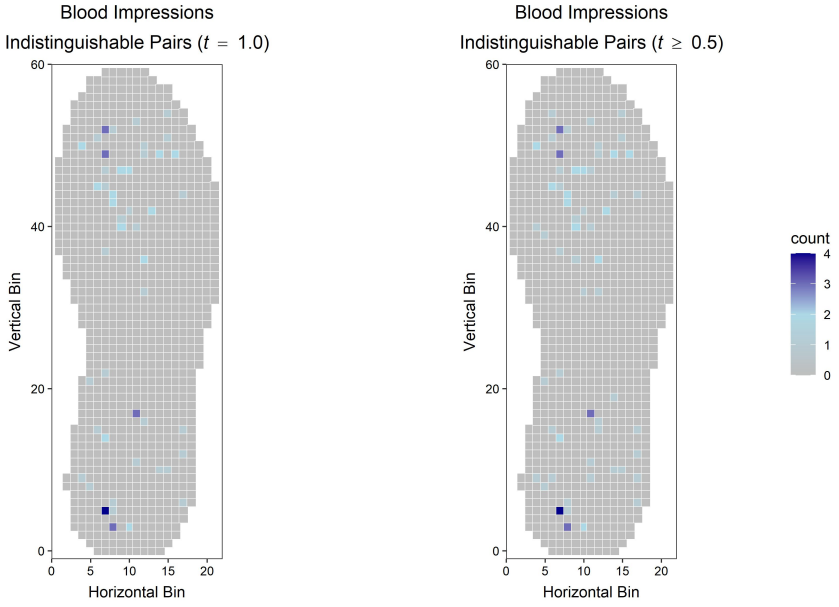
(a) (b)

Figure S7: Heatmaps showing the distribution of indistinguishable pairs across the outsole from the full database of high-quality test impressions for $t = 1.0$ (a) and $t \geq 0.5$ (b) reproduced from Smale and Speir (2023).



(a) (b)

Figure S8: Heatmaps showing the distribution of indistinguishable pairs across the outsole for the test impression subset for $t = 1.0$ (a) and $t \geq 0.5$ (b). Note that these pairs have been plotted using a different density scale than Fig. S7.



(a) (b)

Figure S9: Heatmaps showing the distribution of indistinguishable pairs across the outsole from blood impressions for $t = 1.0$ (a) and $t \geq 0.5$ (b). Note that these pairs have been plotted using a different density scale than Fig. S7.

S8. Footwear Reliability Study (Case 003)

In a footwear reliability study conducted by Richetelli *et al.* (2020), footwear examiners were presented with pairs of questioned and known impressions and tasked with reaching an opinion regarding source association. When comparing the blood impression shown in Fig. S10 to its known mated test impression, nearly 50% of examiners reached a conclusion of high degree of association or identification based on the five circled RACs (and three additional wear patterns not highlighted here; please see original publication for more details). The longest axis of each RAC is reported next to the circle, ranging from approximately 0.7 mm to 4.6 mm in length, with an average ± one standard deviation of 2.8 mm ± 1.5 mm. Note that no claim is being made that examiners based their decision on a single RAC or wear pattern (especially given that this shoe presented at least five RACs and three wear patterns in agreement with its known mate).

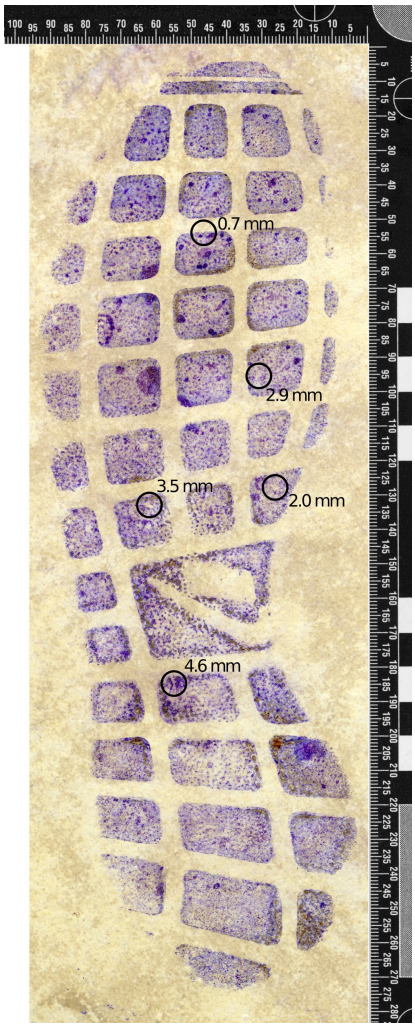


Figure S10: A questioned impression made in blood reproduced from Richetelli *et al.* (2020). The circled RACs (along with three unmarked wear patterns) were used by examiners to reach an opinion of high degree of association or identification when using the SWGTREAD (2013) conclusion standard.

References

* J. A. Speir, N. Richetelli, M. Fagert, M. Hite, W. J. Bodziak, Quantifying randomly acquired characteristics on outsoles in terms of shape and position, Forensic Science International 266 (2016) 399–411. doi:https://doi.org/10.1016/j.forsciint.2016.06.012.
* N. Richetelli, W. J. Bodziak, J. A. Speir, Empirically observed and predicted estimates of chance association: Estimating the chance association of randomly acquired characteristics in footwear comparisons, Forensic Science International 302 (2019) 1–14. doi:https://doi.org/10.1016/j.forsciint.2019.05.049.
* A. N. Smale, J. A. Speir, Estimate of the random match frequency of acquired characteristics in a forensic footwear database, Science & Justice 63 (3) (2023) 427–437. doi:https://doi.org/10.1016/j.scijus.2023.04.007.
* N. Richetelli, L. Hammer, J. A. Speir, Forensic footwear reliability: Part II—Range of conclusions, accuracy, and consensus, Journal of Forensic Sciences 65 (6) (2020) 1871–1882. doi:https://doi.org/10.1111/1556-4029.14551.
* SWGTREAD, Range of Conclusions Standard for Footwear and Tire Impression Examinations, https://www.nist.gov/system/files/documents/2016/10/26/swgtread_10_range_of_conclusions_standard_for_footwear_and_tire_impression_examinations_201303.pdf, 2013 (accessed June 2023).

4. RAC-RMF of Simulated Crime Scene Impressions in Dust

This manuscript and the associated supplemental material have been submitted for publication: A. N. Smale, J. A. Speir, Estimate of the random match frequency of acquired characteristics in footwear: Part II — Simulated crime scene impressions in dust [Manuscript submitted for publication] (2023).

Estimate of the Random Match Frequency of Acquired Characteristics in Footwear: Part II — Simulated Crime Scene Impressions in Dust

Alyssa N. Smale[a], Jacqueline A. Speir[a,*]

[a]*Forensic & Investigative Science, West Virginia University, 208 Oglebay Hall P.O. Box 6121, Morgantown, WV, 26506, United States*

This study serves as Part II of an investigation into the random match frequency of randomly acquired characteristics (RAC-RMF) in footwear evidence. In Part I, RAC-RMF was estimated in a dataset of simulated crime scene impressions deposited in blood. For Part II, a second dataset was created composed of impressions deposited in dust on paper or tile, with the latter lifted using gelatin or Mylar film. A total of 1,513 RACs were identified from more than 160 dust impressions and compared to RACs with positional similarity in test impressions from 1,299 non-mated outsoles. RACs of any size deposited in dust exhibited a 31% decrease in shoes with non-zero RAC-RMFs as compared to their mated test impressions, while those deposited in blood exhibited a 45% increase. When only considering shoes with at least one RAC deemed forensically-reliable (length $\geq$ 2.8 mm), 3.1% of shoes contributing dust impressions and 3.4% of shoes contributing blood impressions exhibited relative RAC-RMFs at a value $\geq$ 0.0008. Although each dataset resulted in a comparable rate for encountering non-zero RAC-RMFs, the estimate for dust was based on twice the number of RAC comparisons (154,477) than those performed when assessing blood (77,566). While these results are considered specific to the non-mated impressions and methods of analysis described herein, and continued work is required before rates can be fully understood and reported in forensic casework, this study encountered non-zero RAC-RMFs for shoes exhibiting at least one forensically-reliable RAC at a more frequent rate than any estimates previously reported.

Keywords: footwear, randomly acquired characteristics, random match frequency, percent area overlap, simulated crime scene impressions, dust impressions

1. Introduction

When evaluating footwear impression evidence, an examiner must assess the similarity of class characteristics and characteristics of use between a questioned impression and a

*Corresponding author

Email address: Jacqueline.Speir@mail.wvu.edu (Jacqueline A. Speir)

known exemplar, as well as the rarity of the characteristics observed. The majority of empirical research has shown that randomly acquired characteristics (RACs) with positional and geometric similarity rarely, if ever, occur on unrelated outsoles [1–5]. However, these studies were performed using relatively small datasets, so the results may be a product of limited samples available for analysis rather than the extremely limited possibility or impossibility of similar RACs occurring on unrelated outsoles. In a recent study [6], random match frequency of randomly acquired characteristics (RAC-RMF) was investigated in the West Virginia University (WVU) footwear database [7], which contains 1,300 shoes of various makes, models, and sizes with over 80,000 RACs, providing a large population for analysis.

In order to estimate RAC-RMF in [6], RACs identified on high-quality test impressions were compared to RACs with positional similarity in test impressions from 1,299 unrelated outsoles in the database. Positional similarity was defined by RACs co-occurring within a shared 5 mm $\times$ 5 mm spatial cell, after mapping to a standardized outsole [7]. This resulted in nearly four million non-mated RAC comparisons [6]. Indistinguishable pairs were detected through a combination of visual comparisons and predictions from a mathematical model based on percent area overlap, as described in [8]. RAC-RMF$_{(m|n \geq 1)}$ was reported for each shoe in the database as a value out of 1,299, representing the number of shoes that shared at least one ($n \geq 1$) indistinguishable RAC and an undetermined number of distinguishable pairs (m) with the held-out shoe, with values as high as 49 out of 1,299 observed [6].

Analysis of a larger database in [6] than used in previous studies showed that it is possible for RAC geometries to repeat in the same relative position on unrelated outsoles. However, the impressions in the WVU database were high-quality test impressions capable of reproducing fine details from the outsole and RACs of very small size. In a casework scenario, questioned impressions are deposited using a variety of media and substrates, typically resulting in lower-quality impressions that vary in totality and clarity. While test impressions provided a useful point of reference for estimating RAC-RMF, further investigation of RAC-RMF from impressions more closely aligned with those received in casework is still needed. To accomplish this, a two-part investigation was designed to estimate RAC-RMF in simulated crime scene impressions.

In Part I of this study, a simulated crime scene impression dataset was created, composed of more than 160 impressions deposited in blood on different tile substrates and enhanced with leucocrystal violet [9]. Each impression was registered to a Handiprint of the corresponding outsole [7] and marked for RACs by a researcher. Although more than 95% of RACs failed to transfer and be detected in blood impressions, at least one RAC was identified in 75% of blood impressions, with a total of 759 RACs across all impressions [9]. In order to estimate RAC-RMF, RACs in blood impressions were compared to RACs with positional similarity in test impressions from 1,299 unrelated outsoles in the WVU database. This resulted in a total of 77,566 non-mated comparisons, which again were accomplished by a combination of visual comparisons and predictions from a mathematical model [8, 9]. In order to provide a more comparable point of reference for these non-mated comparisons than the four million comparisons in the full database of test impressions [6], the known mate of each RAC in a blood impression was identified in the test impression made by the same shoe to form a test impression subset. The 759 RACs in the test impression

subset were compared to RACs with positional similarity in test impressions from 1,299 unrelated outsoles, generating an additional 77,566 comparisons. When comparing these two datasets of equal size, more indistinguishable pairs were observed between the RACs in blood impressions and RACs in non-mated test impressions than between the RACs in the test impression subset and RACs in non-mated test impressions. This was hypothesized to be the result of a reduction in RAC size for features in the blood impressions, thereby decreasing their geometric complexity in binary digitized images, increasing their overall similarity with known non-mates as assessed using the metric of percent area overlap (%A) [8], and affecting the associated RAC-RMF estimates in two ways. First, fewer RACs and non-mated comparisons decreased the magnitude of RAC-RMFs overall. When considering RACs of any size, nearly 70% of shoes had a RAC-RMF$_{(m|n \geq 1)}$ greater than or equal to 1 out of 1,299 in the full database of test impressions [6], while 24% of shoes had a non-zero RAC-RMF in the test impression subset corresponding to the blood impressions [9]. Thus, with fewer RACs available, the chance of finding indistinguishable pairs on unrelated outsoles decreased. Second, the reduction in RAC size caused a nearly two-thirds increase in the number of indistinguishable pairs and a 45% increase in the percentage of non-zero RAC-RMF values for blood impressions compared to the corresponding test impression subset [9]. Overall, shoes with a non-zero RAC-RMF were observed at a rate of 34% for blood impressions when including RACs of any size. Using 2.8 mm as a length threshold based on the average RAC size deemed reliable by examiners in a recent footwear reliability study [10], a conservative rate of 3.4% of shoes possessing at least one RAC of this size had non-zero RAC-RMFs [9].

The results presented in Part I prompted an additional study to determine how findings vary across different media, substrates, and collection techniques. For Part II, a second simulated crime scene impression dataset was created. This dataset was composed of impressions deposited in dust on paper and tile, with the latter lifted using either gelatin or Mylar film and an electrostatic lifter. Using the same general analytical procedures, the dust impressions served as a second lower-quality complement to the high-quality test impressions presented in [6] to better understand the effect of impression quality on RAC transfer and RAC-RMF. In addition, comparison of blood and dust impressions provided insight with regard to the influence of impression type on these same variables.

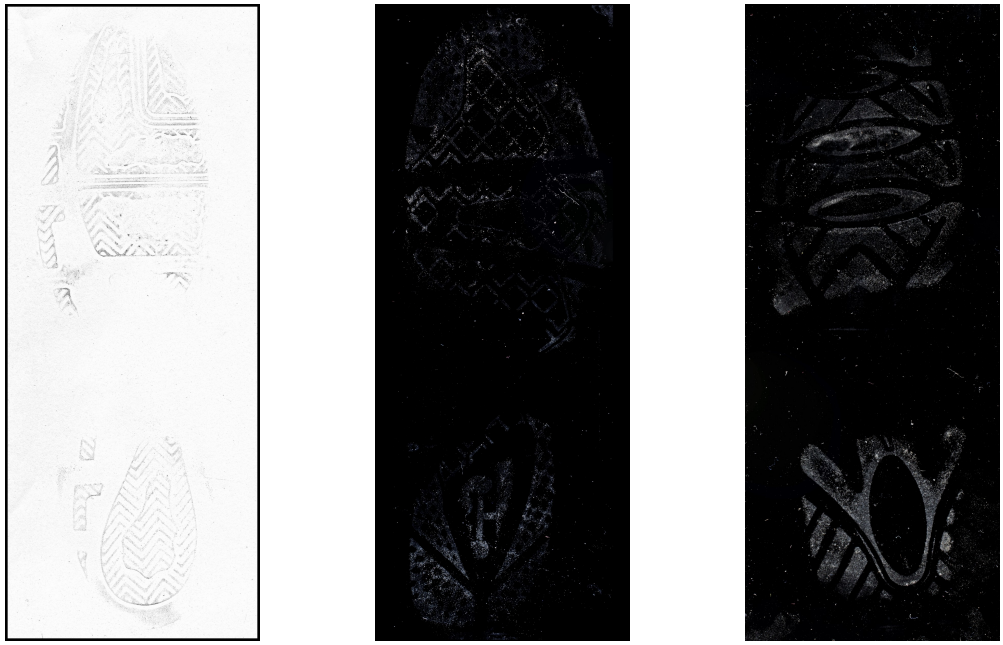
As with previous related studies [6, 9], the results from dust impressions are specific to these shoes and analysis procedures, and therefore do not translate directly to casework estimates. However, the results can be used as a means of calibration for examiners, as the majority of previous research regarding random association in footwear has either been theoretical or shown little-to-no repetition of RAC geometries in empirical studies. In contrast, analysis of the 1,300 shoes in the WVU footwear database has shown that RACs with similar geometries repeat with positional similarity when comparing non-mated test impressions to test impressions [6], blood impressions [9], and, as discussed in this study, dust impressions.

2. Materials and Methods

A subset of shoes from the West Virginia University footwear database [7] was chosen to create a simulated crime scene dataset with impressions deposited in dust. For brevity, the questioned impressions produced from this subset will be referred to as dust impressions. *Note that this subset was different from the one used to create blood impressions and reported on in Part I [9] of this investigation. This choice was purposeful to prevent the need to clean outsoles in between creating questioned impressions, and to avoid the possible transfer and contamination of blood into dust impressions and vice versa. However, in hindsight, this somewhat limits direct comparisons between dust and blood impressions, since each was produced using a different sampling of outsoles.* As with the previous simulated crime scene dataset, shoes with a large number of RACs were selected for the dust dataset. Similar methodology was implemented as described in [9] to analyze these impressions, but with adjustments being made to account for the different impression medium and to attempt to further reduce the possibility of bias in marking. Dataset creation, processing, marking, comparison of RACs, and the evaluation of RAC-RMF are each described separately below.

2.1. Simulated Crime Scene Impressions in Dust

All impressions in this dataset were created using dust particles mixed with a small amount of black fingerprint powder (1-4005 CSI Forensic Supply), as described in [11]. The dust was placed in the bottom of an aluminum tray, and a participant wearing a properly-fitted shoe was instructed to step into the tray and then deposit an impression on white copy paper (PG2014-5 Walmart®) or on the research laboratory floor (made from a tan vinyl tile similar to 54004031 Home Depot®). The impressions deposited on tile were lifted using either a black gelatin adhesive lifter (B-12000 BVDA) or black Mylar film (3054 Evident) with an electrostatic dust lifter (ESP900 Sirchie®) (please see supplemental material section S1 for additional information regarding lifting techniques). The impressions on paper were scanned at a resolution of 600 pixels per inch (PPI) using an Epson® Expression® 11000XL Graphic Arts scanner [11]. The lifted impressions were photographed using a Nikon D7000 camera with a 35 mm lens (66546 Nikkor) at a resolution of 600 PPI in a darkened room using oblique illumination [11]. As necessary, both the scanned images and the photographs of the dust impressions were digitally enhanced using Adobe® Photoshop® to improve contrast. Examples of impressions post-digital enhancement are shown in Fig. 1. There were 55 impressions of each type created (55 on paper, 55 on tile lifted with gelatin, and 55 on tile lifted with Mylar film and an electrostatic lifter), for a total of 165 impressions. For brevity, these impressions are referred to as paper, gelatin, and electrostatic, respectively.



(a) (b) (c)

Figure 1: Examples of dust impressions deposited on paper (a), deposited on vinyl tile and lifted with gelatin (b), and deposited on vinyl tile and lifted with Mylar film and an electrostatic lifter (c).

2.2. Post-Processing

Each enhanced impression was registered to a corresponding Handiprint of its outsole [7]. This was accomplished by selection of eight ground control points and the use of a warping algorithm [7], or by manual registration using Adobe® Photoshop® [9]. Three of the electrostatic-lifted impressions could not be adequately registered to the Handiprint using either method, resulting in 162 useful impressions for further analyses (which was coincidentally the same number of useful blood impressions analyzed in Part I [9]). Of these, 27 were registered using the automated algorithm, while the remaining 135 required manual registration.

2.3. Marking of RACs

Part I [9] of this study described a two-step modified sequential unmasking approach to identify RACs in questioned impressions. This involved the use of a "patch" map that was overlaid on the questioned impression to focus the researcher's attention while searching for RACs. Using this map, possible RACs were marked, but thereafter, the test impression was viewed simultaneously with the simulated crime scene impression to mark additional RACs missed in the first step or to confirm, remove, or adjust the shape of identified RACs (please see [9] for additional information). A drawback of this approach was that any RACs marked or adjusted after viewing the test impression were potentially biased in shape.

In an attempt to further mitigate possible bias when marking the dust impressions, a slightly different approach was implemented that included three separate steps. First, a

patch map was utilized once again to indicate areas in the dust impression where RACs were known to exist on the outsole. Using the transparent patches to focus attention, the researcher simultaneously viewed the marked test impression [7] and the dust impression to search for RACs. The locations of any RACs found in the dust impression were circled, resulting in a more specific location within the transparent patches for each transferred RAC. Two different-colored circles were used during this process; one to indicate that the shape of the RAC located within the circle should be completely filled in, and another to indicate that the outline of the RAC shape should be traced. The reason for this differentiation was to mark each RAC in a manner similar to the way the RACs were marked in the existing test impression, and therefore allow for a direct comparison between the dust impressions and their mated test impressions. A minimum of two months after the circles were added to the patch map, the researcher performed the second step, which involved re-examining the patch maps and associated colored circles. Any possible RACs within these circles were filled in or traced using Adobe® Photoshop®. Finally, the third step was performed, which required the researcher to again simultaneously view the marked test impression with the marked dust impression, and add, confirm, remove, or adjust the shape of identified RACs. Along the way, each RAC was tagged with a label to differentiate between those marked/adjusted before or after viewing the test impressions in the third step.

All marked RACs were extracted to create a binary RAC map for each dust impression, and the shape category of each RAC was determined as described in [8]. Using RAC localization information from the known mated test impression, RACs in dust impressions were likewise localized to the standard outsole [7]. This process ensured that mated RACs in test and dust impressions were localized to the same spatial cell to permit a one-to-one comparison, negating slight shifts due to either the RAC's change in shape resulting in a change in its centroid, and/or slight misalignments during registration.

2.4. Quality Control

Differences in RAC count and shape between the test impressions and the corresponding dust impressions were anticipated. These changes could be due to the impression medium and/or substrate, or researcher variation in marking. In order to differentiate between these factors, a quality control program was implemented to assess researcher variability. A random selection of six dust impressions of each impression type (or approximately 10% of the 165 total impressions) were selected for duplicate marking. These 18 impressions were copied and mixed into the dataset prior to marking so that the researcher was blind to which impressions were being used for quality control. Although variability in detection of RACs was not under evaluation in this study, it was still quantified. Conversely, variability in RAC tracing by the researcher was of greater interest, as this can directly affect the similarity between non-mated RACs, and thus RAC-RMF estimates. To characterize the impact of this, RACs that were marked in both duplicate impressions were compared using percent area overlap (%A) [8], and the effect of researcher variability on probability of indistinguishability was estimated.

2.5. RAC Comparisons

Fig. 2 provides an outline of the comparisons performed between known mated (KM) and known non-mated (KNM) RAC pairs. Arrow colors within this flowchart are referenced moving forward, and any results or abbreviations which have not yet been explained will be discussed in the following sections.

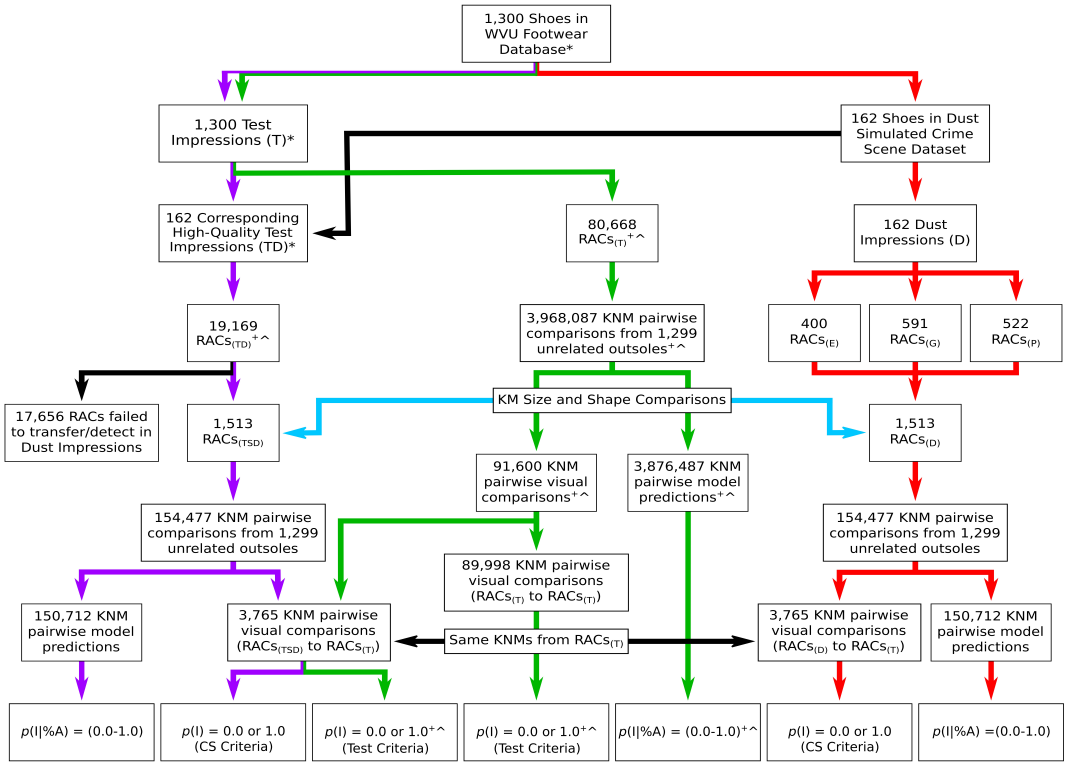


Figure 2: Outline of the RAC comparisons performed in this study and in previous work (*Speir *et al.* (2016) [7], $^{+}$Richetelli *et al.* (2019) [8], ^Smale and Speir (2023) [6]). Please refer to this figure when reading the following sections.

2.5.1. Known Mated RAC Pairs

The mated RAC from the corresponding test impression was identified for each RAC that reproduced in a dust impression, comprising a test impression subset (TSD). Each RAC from the dust impressions ($\text{RAC}_{(\text{D})}$) was compared to its mated RAC in the corresponding test impression subset ($\text{RAC}_{(\text{TSD})}$) to assess any changes in RAC size, shape, and/or categorization as a function of impression quality and type (Fig. 2, blue arrow).

2.5.2. Known Non-Mated RAC Pairs

The procedure for determining non-mated RAC pairs was previously outlined in [6], where each $\text{RAC}_{(\text{T})}$ from a test impression was compared to $\text{RACs}_{(\text{T})}$ with positional similarity (within a 5 mm $\times$ 5 mm spatial cell) from 1,299 test impressions of unrelated outsoles (Fig. 2, green arrows). For this simulated crime scene dataset, each $\text{RAC}_{(\text{D})}$ from dust impressions (Fig. 2, red arrows) and each $\text{RAC}_{(\text{TSD})}$ from the test impression subset (Fig. 2,

purple arrows) was compared to non-mated $RACs_{(T)}$ with positional similarity on the same 1,299 unrelated outsoles.

There were 154,477 non-mated RAC comparisons performed for the dust impression dataset, and another 154,477 non-mated comparisons for the test impression subset. This included 3,765 visual comparisons and 150,712 predictions using the mathematical model described in [8] for $RACs_{(D)}$ versus $RACs_{(T)}$ (Fig. 2, red arrows) and $RACs_{(TSD)}$ versus $RACs_{(T)}$ (Fig. 2, purple arrows), resulting in a total of 7,530 visual comparisons and 301,424 predictions for this study. For the visual comparisons, each RAC pair was assigned a probability of indistinguishability ($p(I)$) of 1.0 if deemed indistinguishable or 0.0 if deemed distinguishable. The indistinguishable pairs included all the RACs visually-identified as indistinguishable in [8] based on variation expected from replicate test impressions (referred to as "Test Criteria" in Fig. 2), along with additional pairs deemed indistinguishable using a more lenient criteria to account for the increased variation expected when comparing crime scene and test impressions (referred to as "CS Criteria" in Fig. 2). For the modeled comparisons, the metric of percent area overlap was used to predict the probability of indistinguishability ($p(I|\%A)$) between 0.0 and 1.0 [8].

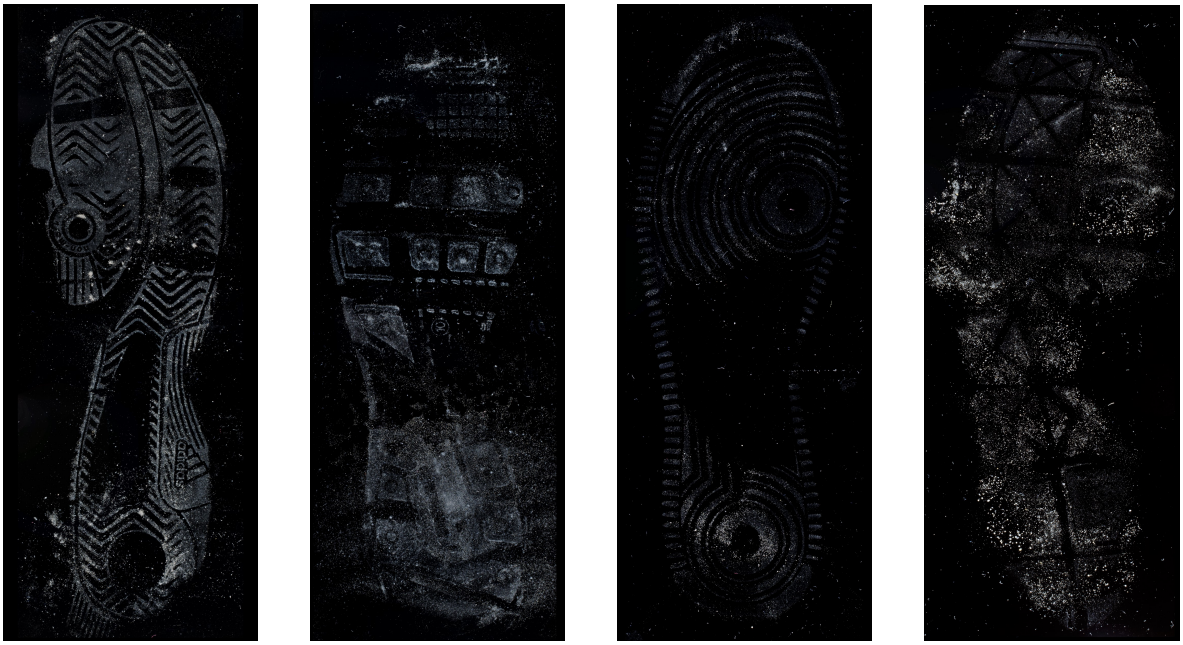
2.6. Random Match Frequency of Randomly Acquired Characteristics

A similar procedure for estimating RAC-RMF as previously described in [6, 9] was implemented for the dust impressions and the test impression subset comprising Part II. Briefly, for any shoe that reproduced at least one RAC in its mated dust impression, $RAC_{(D)}$-$RMF_{(m|n \geq 1)}$ and $RAC_{(TSD)}$-$RMF_{(m|n \geq 1)}$ were computed and reported as the number of shoes out of 1,299 in the full database that shared *at least one* indistinguishable RAC pair ($n \geq 1$) with the held-out shoe, without consideration for the number of distinguishable RACs (m) present. The maximum number of shared indistinguishable RAC pairs (n) between a pair of unrelated outsoles was also determined.

3. Results

3.1. Simulated Crime Scene Impressions in Dust

A total of 165 dust impressions were created, with 55 of each type (electrostatic, gelatin, and paper). Almost all of these impressions reproduced tread elements that spanned toe-to-heel. However, the impressions varied in clarity. There were 114 judged to exhibit a high level of clarity, allowing fine details to be observed across the majority of the impression. All 55 paper impressions were deemed high clarity, along with 36 impressions lifted with black gelatin adhesive and 23 impressions lifted with Mylar film. In contrast, 51 impressions were judged to exhibit excess dust and/or one or more areas of blurred or obscured details. Of these, 32 were lifted using Mylar film and 19 were lifted using gelatin. Figs. 3(a) and 3(b) illustrate examples of electrostatic-lifted impressions with high clarity and areas of excess dust, respectively, and Figs. 3(c) and 3(d) show examples of gelatin-lifted impressions with high clarity and areas of excess dust, respectively.



(a) (b) (c) (d)

Figure 3: Electrostatic-lifted impressions ((a) and (b)) judged to exhibit high clarity and areas of excess dust, respectively, and gelatin-lifted impressions ((c) and (d)) judged to exhibit high clarity and areas of excess dust, respectively.

3.2. Marking of RACs

The 162 impressions that were successfully registered were marked by a single researcher. Of these, 146 (90%) had at least one RAC. For comparison, 122 out of 162 (75%) blood impressions had at least one RAC in Part I of this investigation [9]. However, as previously reported, the subset of shoes used to generate questioned impressions for Part I and Part II of this investigation differed, and the blood dataset in [9] resulted in many more partial impressions than the dust impressions comprising Part II. Marking the circle map on top of the patch map resulted in a total of 1,517 possible RACs from 146 impressions. After viewing the corresponding test impressions, 1,461 RACs were confirmed, 10 RACs were removed after being deemed a pseudo-accidental or failing to generate a one-to-one pair with a RAC in the test impression, another 46 RAC shapes were adjusted, and 6 new RACs were added (please see [9] for additional information). Upon comparison of the marked test and dust impressions, 5 RACs (0.3% of the total) were judged at or near the signal-to-noise limit for reliability, and therefore candidates for removal. However, since none of these RACs contributed to any indistinguishable pairs or non-zero RAC-RMF estimates, their presence in the dataset is considered negligible. The final result was 1,513 $\text{RACs}_{(\text{D})}$ composed of 1,461 unbiased and 52 biased (6 additional + 46 shape changes) shapes. Due to the low/negligible number (3%) of RACs deemed possibly biased, the effect of bias in this dataset was not investigated further.

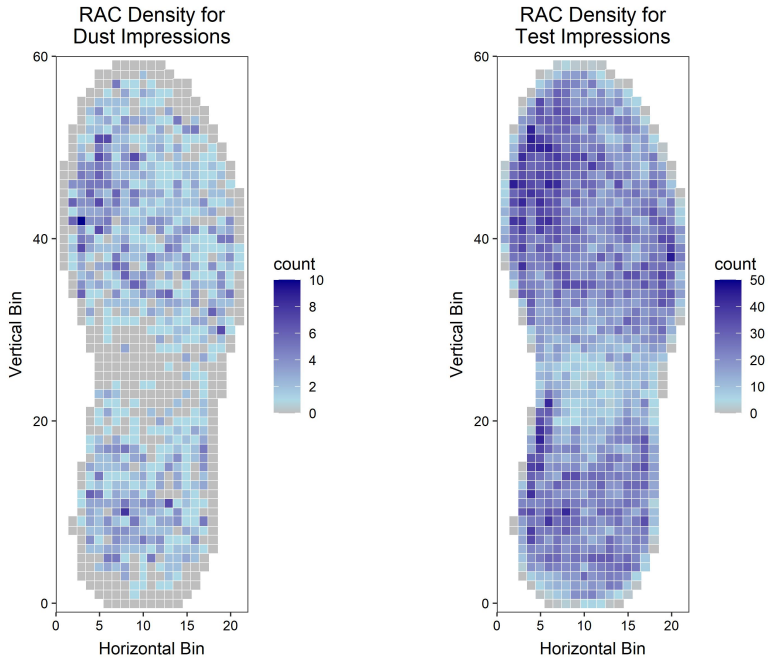
Of the 1,513 $\text{RACs}_{(\text{D})}$ identified and confirmed, 591 were from the gelatin-lifted impressions ($\text{RACs}_{(\text{G})}$), 522 were from the impressions on paper ($\text{RACs}_{(\text{P})}$), and 400 were from

the electrostatic-lifted impressions ($RACs_{(E)}$). Despite having the highest number of RACs (591), gelatin-lifted impressions transferred the lowest proportion of RACs from the corresponding outsoles (6.5%). The distribution of $RACs_{(D)}$ by impression type can be seen in Table 1, along with the RAC count from blood impressions described in Part I [9]. The last two columns of this table report the maximum RAC count for a single impression and the average RAC count with a 95% confidence interval (CI) assuming a Poisson distribution [12]. For all $RACs_{(D)}$, 1,082 (72%) were categorized as variable in shape, while 277 (18%) were compact and 154 (10%) were linear. This general trend for RAC categories was followed for all three dust impression types, as well as the blood impressions [9].

Table 1: RAC transfer results for the three types of dust impressions, along with the combined data for all dust and blood [9] impressions. For the impressions with 1+ RACs and the total RAC count, count is reported first, followed by the percentage in parentheses.

	Impressions		**RACs Transferred**		
	Total	**1+ RACs**	**Total**	**Max.**	**Average, 95% CI**
Electrostatic	52	46 (88.5)	400 (8.8)	71	8, [3, 16]
Gelatin	55	49 (89.1)	591 (6.5)	85	11, [5, 20]
Paper	55	51 (92.7)	522 (9.5)	73	9, [4, 17]
All Dust	162	146 (90.1)	1,513 (7.9)	85	9, [4, 17]
All Blood	162	122 (75.3)	759 (4.0)	30	5, [2, 12]

The spatial cell of each known mate in the test impression was used to localize each RAC in the dust impressions (please see supplemental material section S2 and Fig. S1 for additional information). The 1,513 $RACs_{(D)}$ were found in 645 out of 987 spatial cells with a maximum of 10 RACs per cell, as shown in Fig. 4(a). The corresponding 162 test impressions (TD) had a total of 19,169 $RACs_{(TD)}$, which were distributed in 958 cells with a maximum of 46 RACs per cell, as shown in Fig. 4(b). While RACs were observed across the entire outsole, the highest density was localized to the medial toe in both the dust impressions and their mated test impressions. It can also be seen that fewer RACs from the outsoles' perimeters were reproduced and identified in the dust impressions. This finding was also observed when analyzing the blood impressions in Part I [9], and is again believed to be a function of the difference in collection techniques (walking versus benchtop) [13]. However, the lack of RACs around the perimeter in the dust impressions was not as pronounced as that observed in the blood impressions, which is hypothesized to be the result of a greater percentage of complete impressions in the dust dataset, versus a greater percentage of partial impressions in the blood dataset [9].



(a) (b)

Figure 4: Spatial distribution of RACs across a normalized outsole [7] extracted from the 162 shoes comprising the dust impressions (a) and their corresponding Handiprint test impressions (b). Note that results have been plotted using two different density scales.

The RAC counts in the corresponding 162 test impressions ranged from 27 to 1,328 with an average of 118, 95% CI [98, 141]. There were two shoes that had a RAC count above 400 in the test impressions, which were a left-right pair with 1,196 and 1,328 RACs, respectively. Both shoes were used to create impressions that were lifted with gelatin, resulting in 85 and 81 $\text{RACs}_{(\text{G})}$, respectively. Fig. 5 plots the RAC count in the test impression versus the dust impression for the shoes in this dataset, using different colors to differentiate between the three impression types. There is a weak trend showing that the number of RACs found in a dust impression increases as the number of RACs on the outsole increases. However, there is variability in the number of RACs that reproduced in the dust impressions for shoes with a similar number of RACs in the test impressions, which suggests the influence of other factors, such as impression clarity and quality, on the number of RACs that transfer and are above the signal-to-noise limit for detection in dust impressions. While 90% of the shoes in the dataset had at least one RAC transfer to their mated dust impression, the highest percentage of RACs that transferred from a single outsole was 53.5% (or 23 transferred out of 43 RACs present). In total, approximately 92% of the RACs did not transfer and/or could not be identified in the simulated crime scene impressions deposited in dust, as only 1,513 out of 19,169 RACs were detected.

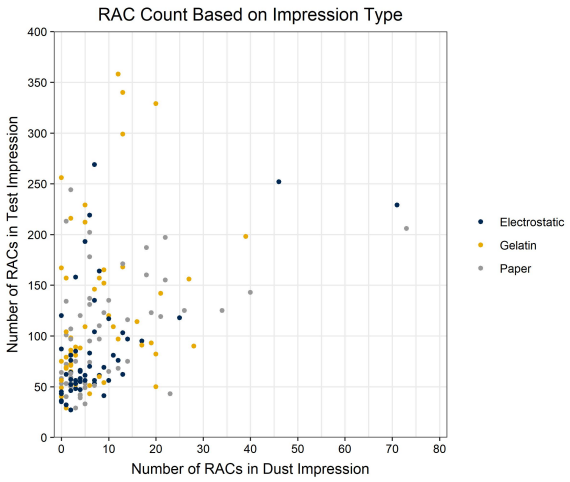


Figure 5: Number of RACs identified in the dust impression versus the number of RACs previously identified in the high-quality test impression for 160 shoes in the simulated crime scene impression dataset. See legend for impression type versus color. Note that to improve visualization for the majority of data points, two shoes with over 1,000 RACs in the test impressions were excluded from this plot (please see supplemental material section S3 and Fig. S2 for a plot of all data).

To investigate RAC size, the longest axis was computed for all 19,169 $\text{RACs}_{(\text{TD})}$ in the 162 corresponding test impressions, the 1,513 $\text{RACs}_{(\text{TSD})}$ from the test impressions that reproduced in the corresponding dust impressions, and the 1,513 $\text{RACs}_{(\text{D})}$ from the dust impressions. Table 2 reports the count and percentage of RACs within different size ranges as well as the maximum RAC length for the dust and blood [9] impressions. The area of each of RAC was also computed. The distribution of areas of the RACs in the test impressions is shown in Fig. 6, where the x-axis represents the inclusive upper limit of each bin (*e.g.*, the first set of bars includes RACs with areas (mm^2) in the range (0.0-0.2] and the second set includes (0.2-0.4]). As observed with the blood impressions [9], larger RACs transferred and were detected more often in dust impressions than smaller RACs.

Table 2: Distribution of RAC sizes as determined by the longest axis for all 19,169 $\text{RACs}_{(\text{TD})}$ in the 162 corresponding test impressions, the 1,513 $\text{RACs}_{(\text{TSD})}$ in the test impression subset, and the 1,513 $\text{RACs}_{(\text{D})}$ in the dust impressions†. For reference, the corresponding results for blood impressions‡ are reproduced from [9]. Count is reported first, followed by the percentage in parentheses.

	Count and Percentage (%) of RACs						**Maximum**
	(0–1] mm	**(1–2] mm**	**(2–3] mm**	**(3–4] mm**	**(4–5] mm**	**>5 mm**	**Length (mm)**
$^{\dagger}19{,}169\ \text{RACs}_{(\text{TD})}$	5,889 (30.7)	6,313 (32.9)	3,020 (15.8)	1,463 (7.6)	809 (4.2)	1,675 (8.7)	71.7
$^{\dagger}1{,}513\ \text{RACs}_{(\text{TSD})}$	318 (21.0)	519 (34.3)	285 (18.8)	115 (7.6)	84 (5.6)	192 (12.7)	46.7
$^{\dagger}1{,}513\ \text{RACs}_{(\text{D})}$	243 (16.1)	589 (38.9)	321 (21.2)	150 (9.9)	79 (5.2)	131 (8.7)	36.7
$^{\ddagger}18{,}887\ \text{RACs}_{(\text{TB})}$	5,792 (30.7)	5,875 (31.1)	3,036 (16.1)	1,660 (8.8)	987 (5.2)	1,537 (8.1)	42.2
$^{\ddagger}759\ \text{RACs}_{(\text{TSB})}$	139 (18.3)	228 (30.0)	133 (17.5)	90 (11.9)	78 (10.3)	91 (12.0)	25.2
$^{\ddagger}759\ \text{RACs}_{(\text{B})}$	208 (27.4)	307 (40.4)	125 (16.5)	62 (8.2)	35 (4.6)	22 (2.9)	9.1

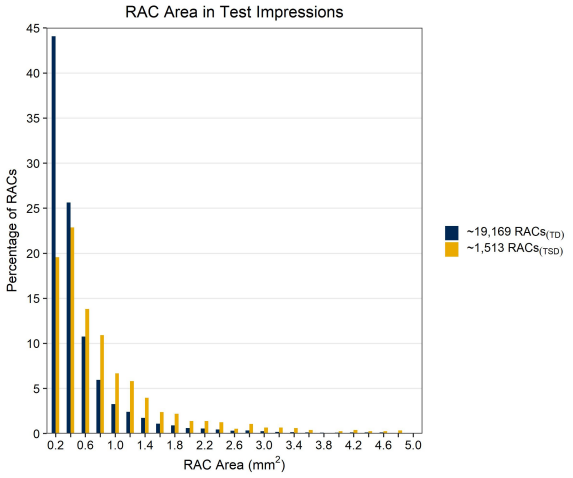


Figure 6: Distribution of RAC areas in test impressions from the 162 shoes used to create dust impressions. The blue-black distribution reports the area of ~19,169 $RACs_{(TD)}$ on these shoes, while the gold distribution reports the area of ~1,513 $RACs_{(TSD)}$ in the test impression subset. Only RACs with an area less than or equal to 5.0 mm^2 are plotted, but this includes more than 98% of all RACs under investigation.

3.3. Known Mated RAC Pairs

$RACs_{(D)}$ and $RACs_{(TSD)}$ were compared to evaluate the effect of impression type on their size and shape (Fig. 2, blue arrow). Table 2 reported the distribution of the longest axis for $RACs_{(D)}$ in the dust impressions and the known mated $RACs_{(TSD)}$ in test impressions. A plot of the relative areas of each of these RACs is shown in Fig. 7. Overall, approximately 57% of RACs were larger in the dust impression than in the corresponding test impression, while 42% were smaller and 6 RACs (<0.5%) were the same size in both impressions. These results are in contrast to that observed for the simulated crime scene impressions prepared in Part I, where 66% of RACs were *smaller* in the blood impressions than the test impressions [9]. Table 3 shows the relative proportion of RACs that were larger in the simulated crime scene or test impressions for the three different types of dust impressions, along with the proportions for all dust and blood [9] impressions. Although not quite a 50:50 split, impressions on paper and impressions lifted with Mylar film had similar proportions of RACs that were larger in the dust impression and RACs that were larger in the test impression. However, approximately 67% of the RACs from impressions lifted with gelatin were larger than their test impression mates, suggesting that gelatin-lifted impressions are driving any perception of increased size in dust impressions overall. This finding is further confirmed in Fig. 7, where more gold data points are observed below the equivalency line.

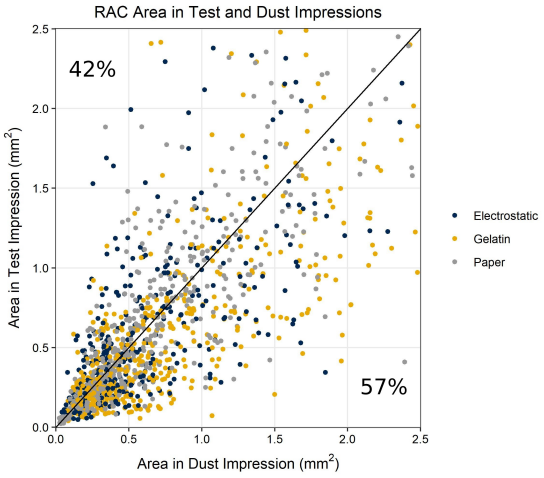


Figure 7: Relative area of 1,370 (91%) of the 1,513 known mated RACs in test and dust impressions (please see supplemental material section S4 and Fig. S3 for plot of all data including RAC areas greater than 2.5 mm^2). The diagonal line represents no change in RAC size between impression types.

Table 3: Proportion of RACs that were larger in the simulated crime scene (SCS) impression or test impression for the three types of dust impressions, along with the results for all dust and blood [9] impressions.

		SCS > Test		**SCS = Test**		**Test > SCS**	
	RAC Count	**Count**	**%**	**Count**	**%**	**Count**	**%**
Electrostatic	400	206	51.5	0	0	194	48.5
Gelatin	591	395	66.8	2	0.3	194	32.8
Paper	522	267	51.1	4	0.8	251	48.1
All Dust	1,513	868	57.4	6	0.4	639	42.2
All Blood	759	255	33.6	3	0.4	501	66.0

In addition to a change in size, the shapes of some RACs were also affected when reproduced in a dust impression. There were 1,203 RACs (80%) that were classified in the same shape category as their mated RAC in the high-quality test impression. The remaining 310 experienced a shape change that was substantial enough to be classified into a different shape category. The confusion matrix shown in Table 4 compares the categories of these RACs in dust and test impressions. As a point of reference, 76% of RACs maintained their original classification from the test impression for the blood impression dataset analyzed in Part I [9].

Table 4: Confusion matrix reporting the original categories of the $RACs_{(TSD)}$ in the test impressions and the categories of the same $RACs_{(D)}$ in the dust impressions.

		Dust Impression		
		Linear	**Compact**	**Variable**
Test Impression	**Linear**	103	16	56
	Compact	5	208	134
	Variable	46	53	892

3.4. Quality Control

Differences in RAC count between the 18 impressions that were marked twice for quality control were computed (please see supplemental material section S5 and Table S1 for detailed results). The primary goal of repeated markings for the impressions that were part of the quality control study was to assess the variation in the researcher's tracing judgments when marking RAC shapes. When comparing duplicate impressions, a total of 170 RAC pairs were found to repeat in both mark-ups. Of the 340 RACs in these pairs, 330 were unbiased and 10 were labeled possibly biased. The minimum percent area overlap value of these 170 pairs was 26.1%, the maximum was 95.5%, and the average ± one standard deviation was 72.6% ± 16.7%, indicating that tracing judgment varies on average by approximately 30%. This was also observed for the quality control set associated with the blood impressions analyzed in Part I [9]. As previously investigated, a $p(\mathrm{I}|\Delta 30\%)$ in the range for non-mated pairs is expected to generate a change in the probability of indistinguishability that is less than 0.08 [9], and therefore the researcher's variability in marking RACs in the dust impressions represents a negligible contribution to uncertainty in RAC-RMF estimates.

3.5. Known Non-Mated RAC Pairs

For a given RAC, its non-mated pairs were defined as $RACs_{(T)}$ with positional similarity in test impressions identified from 1,299 unrelated outsoles. The $RACs_{(D)}$ from the dust impressions and their known mated $RACs_{(TSD)}$ in the test impression subset were compared to the same non-mated $RACs_{(T)}$ in the full WVU database [7]. Results of these comparisons for the dust impressions (Fig. 2, red arrows) are discussed in detail, while the results of the test impression subset (Fig. 2, purple arrows) and the full database of test impressions [6] (Fig. 2, green arrows) are provided for comparison.

These non-mated comparisons resulted in 154,477 pairwise RAC comparisons for the dust dataset and another 154,477 for the corresponding test impression subset. For each dataset, 3,765 of the total 154,477 RAC comparisons were performed visually by two researchers and 150,712 were predictions made by the mathematical model [8]. For the visual comparisons of $RACs_{(TSD)}$ and $RACs_{(T)}$, 172 pairs were deemed indistinguishable, and thus assigned a $p(\mathrm{I})$ value of 1.0. Forty-four of these pairs were deemed indistinguishable based

on visual comparisons in [8] and a criteria which only permitted variation expected in replicate test impressions. The additional 128 pairs were deemed indistinguishable due to the more permissive criteria which allowed for variation expected when comparing a questioned and a known impression (please see supplemental material section S6 for a discussion on uncertainty in these pairs, and section S7 and Figs. S4 and S5 for examples). For the visual comparisons of $\text{RACs}_{(\text{D})}$ and $\text{RACs}_{(\text{T})}$, 97 pairs were deemed indistinguishable.

Any indistinguishable RAC pairs confirmed by visual comparison were considered *certain* indistinguishable pairs at a threshold of $t = 1.0$. Lower thresholds of $t \geq 0.75$ and $t \geq 0.5$ were used to define *plausible* and *possible* pairs, respectively, as detected by the mathematical model. For both datasets, there were no additional indistinguishable pairs detected at $t \geq 0.75$, so this threshold will not be discussed further. At a threshold of $t \geq 0.5$, three additional *possible* pairs were detected for the test impression subset and four were detected for the dust impression dataset. All probability values resulting from the combined visual comparisons and predictions from the mathematical model [8] are reported in Table 5.

Table 5: Distribution of probability values resulting from all comparisons, including $\text{RACs}_{(\text{T})}$ versus $\text{RACs}_{(\text{T})}$ (test impressions) [6], $\text{RACs}_{(\text{TSD})}$ versus $\text{RACs}_{(\text{T})}$ (test impression subset), and $\text{RACs}_{(\text{D})}$ versus $\text{RACs}_{(\text{T})}$ (dust impressions).

	Test Impressions		**Test Impression Subset**		**Dust Impressions**	
Probability	**Count**	**%**	**Count**	**%**	**Count**	**%**
{0.0}	89,419	2.25	3,593	2.33	3,668	2.37
(0.0-0.25)	3,875,532	97.7	150,683	97.5	150,688	97.5
[0.25-0.5)	800	0.0202	26	0.0168	20	0.0129
[0.5-0.75)	133	0.00335	3	0.00194	4	0.00259
[0.75-1.0)	22	0.000554	0	0	0	0
{1.0}	2,181	0.0550	172	0.111	97	0.0628
Total Comparisons	3,968,087 (Fig. 2, green arrows)	—	154,477 (Fig. 2, purple arrows)	—	154,477 (Fig. 2, red arrows)	—

The longest axis of each RAC in an indistinguishable pair was further evaluated. The RACs in the 101 *possible* indistinguishable pairs ($t \geq 0.5$) for the dust impression dataset ($\text{RACs}_{(\text{D})}$ + $\text{RACs}_{(\text{T})}$) had a range in length of approximately 0.4 mm to 6.2 mm with a median value of 0.9 mm, while the RACs in the 175 pairs from the test impression subset ($\text{RACs}_{(\text{TSD})}$ + $\text{RACs}_{(\text{T})}$) had a range in length of approximately 0.4 mm to 7.2 mm with a median value of 0.8 mm. For reference, non-mated $\text{RACs}_{(\text{T})}$ in the 2,336 pairs from the full database of test impressions ranged from approximately 0.2 mm to 9.4 mm with a median value of 0.9 mm [6]. The longest axis measurements for approximately 99% of these RACs are shown in Fig. 8, where the x-axis indicates the inclusive upper limit of each bin. Selecting 2.0 mm as the longest axis measurement beyond which the majority of RACs (all RACs in the dust impression pairs, 82% in the test impression subset pairs, and 98% in the test impression pairs) were linear features, the cumulative percentage of RACs greater than or equal to 2.0 mm was 15.0% for the dust impressions, 16.9% for the test impression

subset, and 17.8% for all test impressions. As a point of reference, 14.3% of the RACs in the indistinguishable pairs for the blood impressions and 22.0% for the indistinguishable pairs in the corresponding test impression subset were greater than or equal to 2.0 mm [9].

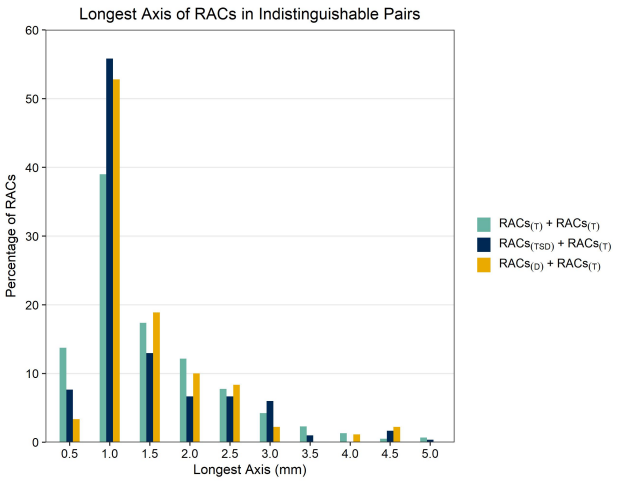


Figure 8: Distribution of longest axis measurements of RACs in *possible* indistinguishable pairs ($t \geq 0.5$) for the full database of test impressions ($\text{RACs}_{(\text{T})}$ + $\text{RACs}_{(\text{T})}$) [6], the test impression subset ($\text{RACs}_{(\text{TSD})}$ + $\text{RACs}_{(\text{T})}$), and the dust impressions ($\text{RACs}_{(\text{D})}$ + $\text{RACs}_{(\text{T})}$). Only RACs shorter than or equal to 5.0 mm are plotted, but this includes approximately 99% of the RACs in indistinguishable pairs across all three datasets.

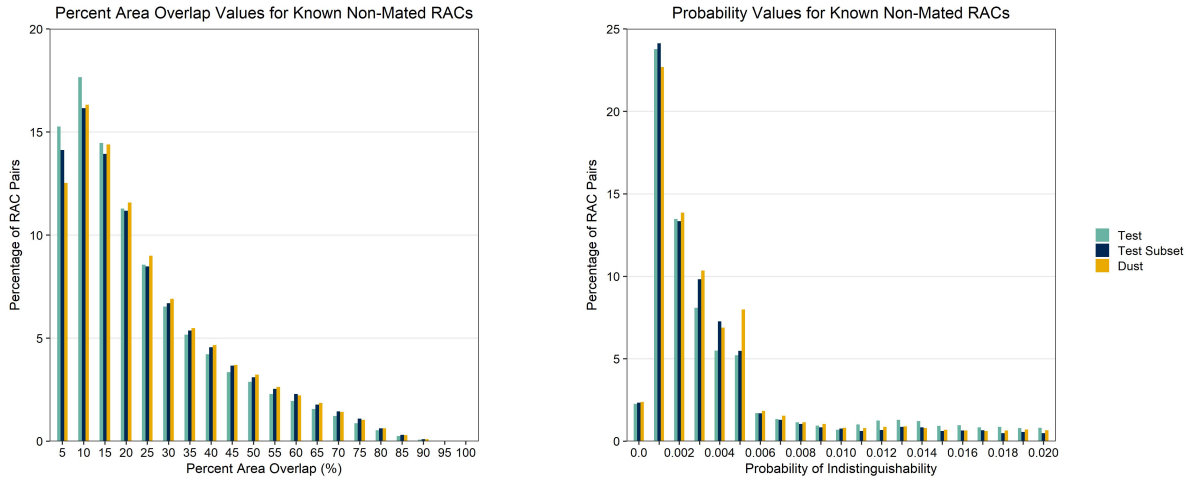
For the 97 pairs deemed indistinguishable by visual comparison in the dust impression dataset, the minimum percent area overlap value was 22.1%, the maximum was 92.2%, and the average ± one standard deviation was 72.8% ± 17.1%. This wide range highlights some of the limitations associated with using a percent area overlap similarity metric, as even some pairs with a low %A value were deemed indistinguishable by a human observer, as further explained in [6]. The additional pairs detected by the mathematical model at $t \geq 0.5$ ranged from 74.5% to 92.3%, with an average ± one standard deviation of 80.9% ± 7.85%. The percent area overlap values for all three datasets are reported in Table 6. When comparing the three datasets, the %A values are similar overall for each threshold.

Table 6: Percent area overlap values of indistinguishable pairs for the full database of test impressions [6] (left), the test impression subset (center), and dust impressions (right) for each threshold.

				Percent Area Overlap (%)								
Probability	**RAC Pairs**			**Minimum**			**Maximum**			**Average ± 1 SD**		
$t = 1.0$	2,181	172	97	12.3	25.7	22.1	100	93.8	92.2	71.5 ± 16.0	74.9 ± 16.2	72.8 ± 17.1
$1.0 > t \geq 0.5$	155	3	4	74.4	75.2	74.5	97.9	78.2	92.3	79.9 ± 5.67	77.0 ± 1.60	80.9 ± 7.85

The distributions of percent area overlap values and associated probabilities of indistinguishability for all three datasets are shown in Fig. 9. All %A values are displayed in Fig. 9(a), but only those probabilities less than 0.02 (≈ 75%) are plotted in Fig. 9(b).

In both plots, the x-axis indicates the inclusive upper limit of each bin. These plots reveal a similar trend between all three datasets. The test impression subset and the dust impression dataset exhibit a shift toward higher percent area overlap values than the full database of test impressions [6]. The dust impressions have higher probabilities than either of the test impression datasets in the range of 0.002-0.005, but this trend is not as pronounced for higher probability values. Overall, this data suggests that $\text{RACs}_{(\text{D})}$, $\text{RACs}_{(\text{TSD})}$, and $\text{RACs}_{(\text{T})}$ all produce similar probability of indistinguishability values when compared to non-mated $\text{RACs}_{(\text{T})}$. This is in contrast to the blood impression dataset investigated in Part I [9], where it was found that blood impressions exhibited a noticeable shift toward higher percent area overlap and probability of indistinguishability values compared to the test impression subset and the full database of test impressions, indicating that $\text{RACs}_{(\text{B})}$ were more similar than either $\text{RACs}_{(\text{TSB})}$ or $\text{RACs}_{(\text{T})}$ to non-mated $\text{RACs}_{(\text{T})}$.



(a) (b)

Figure 9: Distribution of percent area overlap (a) and probability of indistinguishability (b) values for known non-mated RAC pairs from different impression datasets.

RAC category was also explored for these indistinguishable RAC pairs. Table 7 shows the percentage of RAC pairs of each category for all three datasets. Linear pairs were the most common for the full database of test impressions, test impression subset, and dust impressions at $t \geq 0.5$, as well as the full database of test impressions at $t = 1.0$. However, compact pairs were the most common for the test impression subset and dust impressions at $t = 1.0$. For all three datasets at $t = 1.0$, mixed pairs were slightly more common than variable pairs. For the full database of test impressions at $t \geq 0.5$, mixed pairs were the second-most common, followed by compact and then variable. Due to the low number of pairs for the test impression subset and dust impressions at $t \geq 0.5$, it is difficult to discern a trend between the remaining three categories.

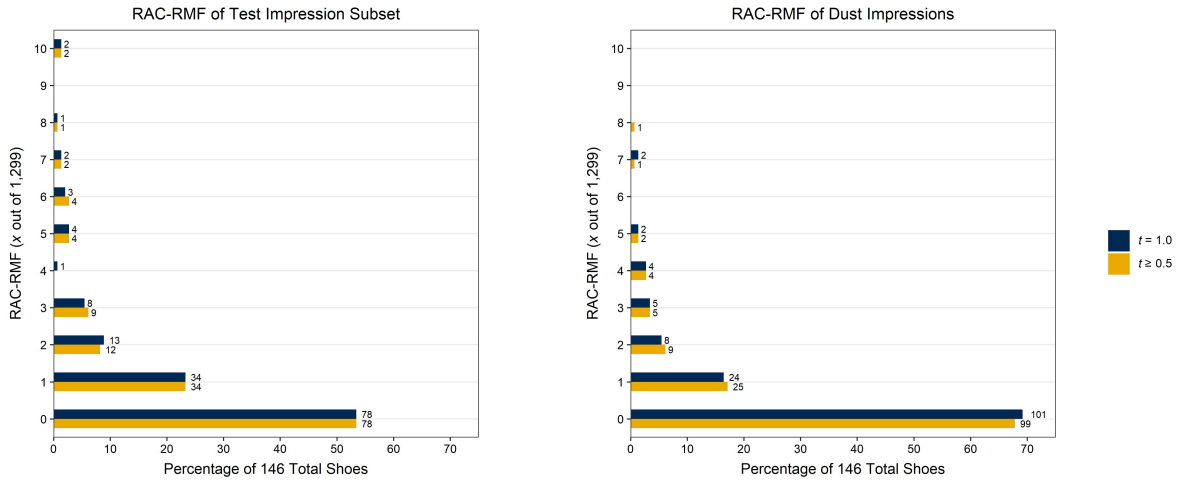
Table 7: Percentage of indistinguishable pairs of each RAC category for all test impressions (left) [6], the test impression subset (center), and dust impressions (right) for each threshold.

Probability	RAC Pairs			Percentage of Total Pairs (%) Linear			Compact			Variable			Mixed		
$t = 1.0$	2,181	172	97	55.6	31.3	30.9	39.0	57.6	45.4	1.7	4.1	11.3	3.7	7.0	12.4
$1.0 > t \geq 0.5$	155	3	4	60.6	66.7	75.0	7.1	0	0	0	0	25.0	32.3	33.3	0

Lastly, the distribution of the indistinguishable pairs across the standardized outsole [7] was determined. The 97 visually-confirmed indistinguishable pairs in the dust impression dataset were found in 71 different cells, with a maximum of four pairs in a single cell. The four additional indistinguishable pairs detected by the mathematical model were found in four different cells, only one of which was the same as the location of an indistinguishable pair with a probability of indistinguishability equal to 1.0. These 101 pairs were distributed across the outsole, but may exhibit a weak degree of clustering (please see supplemental material section S8 and Figs. S6–S8 for additional information). However, spatial autocorrelation was not computed due to the small sample size.

3.6. Random Match Frequency of Randomly Acquired Characteristics

There were 146 dust impressions that had at least one RAC. RAC-RMF$_{(m|n \geq 1)}$ was computed for each of the 146 shoes that created these impressions by determining the number of unrelated test impressions out of 1,299 that shared at least one indistinguishable RAC with the held-out shoe. Bar plots of the results are shown in Fig. 10 for the test impression subset (10(a)) and the dust impressions (10(b)), where RAC-RMF$_{(m|n \geq 1)}$ is displayed as a value out of 1,299 on the y-axis, and the number of shoes corresponding to each RAC-RMF value is shown both as a percentage out of 146 on the x-axis and as a count to the right of each bar. For the dust impressions at $t = 1.0$, 101 shoes (69.2%) had a RAC$_{(\mathrm{D})}$-RMF$_{(m|n \geq 1)}$ of 0 out of 1,299, meaning that these shoes did not share an indistinguishable pair with any of the 1,299 unrelated outsoles. The maximum RAC$_{(\mathrm{D})}$-RMF$_{(m|n \geq 1)}$ was 7 out of 1,299 for two shoes (1.4%). At $t \geq 0.5$, 99 (67.8%) shoes had a RAC$_{(\mathrm{D})}$-RMF$_{(m|n \geq 1)}$ of 0 out of 1,299 and the maximum value increased to 8 out of 1,299 for a single shoe (0.7%). For the test impression subset at both thresholds, 78 shoes (53.4%) had a RAC$_{(\mathrm{TSD})}$-RMF$_{(m|n \geq 1)}$ of 0 out of 1,299, and the maximum RAC$_{(\mathrm{TSD})}$-RMF$_{(m|n \geq 1)}$ observed was 10 out of 1,299 for two shoes (1.4%).



(a) (b)

Figure 10: RAC-RMF$_{(m|n \geq 1)}$ as a value out of 1,299 for $t = 1.0$ and $t \geq 0.5$ for the test impression subset (a) and the dust impressions (b). As an example, 8 out of 146 (5.5%) dust impressions shared at least one indistinguishable pair with 2 out of 1,299 unrelated outsoles in the full database at $t = 1.0$.

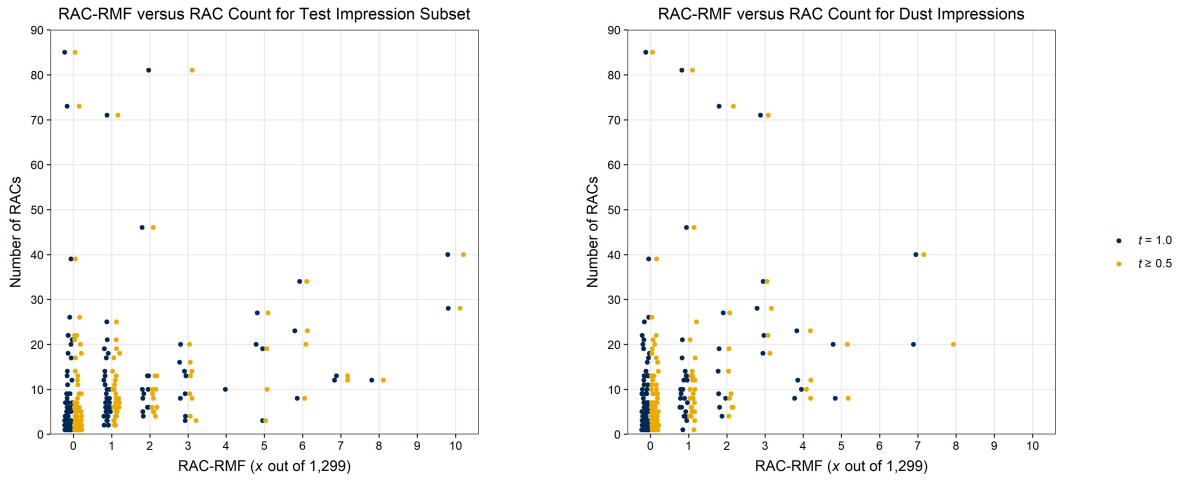
Table 8 shows the distribution of RAC-RMF$_{(m|n \geq 1)}$ values by impression type at $t = 1.0$ and $t \geq 0.5$. Impressions on paper had the highest percentage of impressions with a RAC$_{(D)}$-RMF$_{(m|n \geq 1)}$ value greater than or equal to 1 out of 1,299 at both thresholds. This trend may be explained by the number of impressions with at least one RAC, which was highest for paper and lowest for electrostatic-lifted impressions. In addition, impressions deposited on paper were judged to be of higher clarity than the other two impression types. The impression with the maximum RAC$_{(D)}$-RMF$_{(m|n \geq 1)}$ value observed (8 out of 1,299 at $t \geq 0.5$) was lifted using gelatin, which may be explained by gelatin-lifted impressions having the highest total number of RACs, and therefore a greater chance of finding at least one indistinguishable RAC on an unrelated shoe. However, these differences between impression types are unlikely to be significant, although statistical test was not performed.

Table 8: Distribution of non-zero and maximum RAC-RMF$_{(m|n \geq 1)}$ values by impression type for dust and blood [9] impressions at $t = 1.0/t \geq 0.5$.

	Impressions with 1+ RACs	**Total RAC Count**	**$\geq$ 1 out of 1,299 Count**	**%**	**Max. RAC-RMF (out of 1,299)**
Electrostatic	46	400	10/11	21.7/23.9	4/4
Gelatin	49	591	15/15	30.6/30.6	7/8
Paper	51	522	20/20	39.2/39.2	7/7
All Dust	146	1,513	45/47	30.8/32.2	7/8
All Blood	122	759	37/42	30.3/34.4	6/6

RAC-RMF$_{(m|n \geq 1)}$ as a function of RAC count per shoe was also investigated for the test impression subset (Fig. 11(a)) and the dust impressions (Fig. 11(b)) for $t = 1.0$ and t

≥ 0.5. As previously observed with the full database of test impressions [6] and the blood impressions investigated in Part I [9], there is a slight upward trend showing that RAC-RMF$_{(m|n \geq 1)}$ increases with RAC count. However, the shoes with the highest RAC count (between 70 and 85 RACs) all have RAC-RMF$_{(m|n \geq 1)}$ values less than or equal to 3 out of 1,299. As a result, the upward trend is judged to be less pronounced in the dust impression dataset than the other datasets [6, 9].



(a) (b)

Figure 11: RAC-RMF$_{(m|n \geq 1)}$ as a function of RAC count for the RACs$_{(\mathrm{TSD})}$ in the test impression subset (a) and RACs$_{(\mathrm{D})}$ in the dust impressions (b) for $t = 1.0$ and $t \geq 0.5$. The thresholds were dodged left and right by 0.5, and data points within each threshold were jittered horizontally $\pm$ 0.4 to improve visualization.

Lastly, all pairs of shoes with at least one shared indistinguishable RAC pair were evaluated to determine the maximum number of indistinguishable RAC pairs between unrelated outsoles. As with the blood impressions [9], no more than one shared indistinguishable RAC pair (n =1) was ever observed. This result could be a function of the decreased number of RACs available in these impressions compared to the full database of test impressions, where values of $n = 3$ and $n = 5$ were observed for $t = 1.0$ and $t \geq 0.5$, respectively [6].

3.6.1. Forensic Implications

In Part I [9] of this investigation, the size of RACs in blood impressions was evaluated in the context of a footwear reliability study [10]. The goal was to determine a reasonable threshold for the size/length/area of a RAC in a questioned impression that is deemed valuable in a comparison when forming an opinion regarding possible source associations/disassociations. Extrapolation of results from [10], and based on the reasoning described in [9], this average threshold $\pm$ one standard deviation was found to be 2.8 mm $\pm$ 1.5 mm, with a minimum of 0.7 mm. Unfortunately, this same reliability study [10] did not allow for a determination of this threshold specific to questioned dust impressions. Therefore, in the absence of any other data to inform expectations, the length of RACs in the blood impression marked by examiners in [10] was re-used to provide a point of reference.

Approximately 95% of the RACs in the dust impressions and 90% of the RACs in the test impression subset were larger than 0.7 mm, which was the smallest RAC identified by examiners in the questioned impression reviewed in [10]. When compared to the average RAC length, 400 (26%) of the RACs detected in the dust impressions and 425 (28%) of the RACs in the test impression subset were longer than 2.8 mm. Of the 400 RACs in the dust impressions, 180 were from gelatin-lifted impressions, 139 were from impressions on paper, and 81 were from electrostatic-lifted impressions. If only considering RACs with length greater than or equal to 2.8 mm, there would be four indistinguishable pairs in the dust impression dataset and nine in the test impression subset, all of which were visually-confirmed ($t = 1.0$). This translates to RAC-RMF$_{(m|n \geq 1)}$ values of at least 1 out of 1,299 for 3 (3.1%) out of 98 shoes in the dust impression dataset and 8 (8.2%) out of 98 in the test impression subset, where 98 is the number of shoes with at least one RAC longer than or equal to 2.8 mm in the test impression subset. For the three shoes in the dust impression dataset with a non-zero RAC-RMF, two corresponded to an impression lifted with gelatin and one corresponded to an impression on paper. In summary, Table 9 compares the proportion of shoes with a non-zero RAC-RMF for features at least 2.8 mm in length for dust, blood [9], and test [6] impressions. For the blood dataset, 3 (3.4%) and 7 (7.9%) out of 89 shoes with at least one RAC longer than or equal to 2.8 mm had non-zero RAC-RMFs in the blood impressions and the corresponding test impression subset, respectively [9]. Using the same length threshold, 102 (9.0%) out of 1,131 shoes had a non-zero RAC-RMF for the full database of test impressions [6]. The proportions of shoes with non-zero RAC-RMFs when considering RACs of any size are also included in this table for comparison.

Table 9: Percentage of shoes with non-zero RAC-RMFs including only those RACs with a length greater than or equal to 2.8 mm at $t = 1.0$ versus RACs of any length/size at $t \geq 0.5$ for the dust†, blood‡ [9], and full database of test$^{\wedge}$ [6] impressions. Shoe count is reported as a fraction of the number of shoes with at least one RAC meeting the RAC length criteria (based on the test impression subset for the first four comparisons), followed by the percentage in parentheses. For the simulated crime scene (SCS) datasets, the percentage change was calculated with reference to the associated test impression subset (TS(D or B)) using (SCS – TS)/TS × 100%, where a negative value indicates a decrease (*e.g.*, using the first two rows as an example, 8 versus 3 shoes is a 63% decrease (3 – 8)/8 × 100% = -63%).

Comparison	**Probability**	**RAC Length**	**Shoes with RAC-RMF ≥ 1 out of 1,299**	**Percentage Change**
†RACs$_{(TSD)}$ versus RACs$_{(T)}$	$t = 1.0$	≥ 2.8 mm	8/98 (8.2%)	-63%
†RACs$_{(D)}$ versus RACs$_{(T)}$	$t = 1.0$	≥ 2.8 mm	3/98 (3.1%)	
†RACs$_{(TSD)}$ versus RACs$_{(T)}$	$t \geq 0.5$	Any	68/146 (46.6%)	-31%
†RACs$_{(D)}$ versus RACs$_{(T)}$	$t \geq 0.5$	Any	47/146 (32.2%)	
‡RACs$_{(TSB)}$ versus RACs$_{(T)}$	$t = 1.0$	≥ 2.8 mm	7/89 (7.9%)	-57%
‡RACs$_{(B)}$ versus RACs$_{(T)}$	$t = 1.0$	≥ 2.8 mm	3/89 (3.4%)	
‡RACs$_{(TSB)}$ versus RACs$_{(T)}$	$t \geq 0.5$	Any	29/122 (23.8%)	45%
‡RACs$_{(B)}$ versus RACs$_{(T)}$	$t \geq 0.5$	Any	42/122 (34.4%)	
$^{\wedge}$RACs$_{(T)}$ versus RACs$_{(T)}$	$t = 1.0$	≥ 2.8 mm	102/1,131 (9.0%)	—
$^{\wedge}$RACs$_{(T)}$ versus RACs$_{(T)}$	$t \geq 0.5$	Any	890/1,300 (68.5%)	—

3.7. Summary and Comparison

While there are likely several confounding factors impacting the results obtained from test [6], blood [9], and dust impressions, three trends were observed to be associated with RAC-RMF, including dataset size, the medium/substrate/quality/clarity/totality of the simulated crime scene impressions, and lastly, RAC size.

3.7.1. Dataset Size

Comparison of the full database of test impressions [6] and the test impression subsets from Part I (TSB) [9] and Part II (TSD) allowed for investigation into the effect of dataset size on RAC-RMF. These datasets contained test impressions of the same quality, as TSB and TSD are subsets of the larger database. However, the full database of test impressions contained many more RACs than either test impression subset. In addition, only subsamples of the RACs in the test impressions were evaluated based on the number of RACs in the associated simulated crime scene impressions (759 RACs in TSB and 1,513 RACs in TSD). Table 10 shows a comparison of the three datasets, where a horizontal line is used to differentiate between subsampling steps. In order to compare the influence of dataset size on RAC-RMF, the number of RACs in each dataset with a major axis greater than or equal to 2.8 mm was determined, and the number of shoes with non-zero RAC-RMF values using only these RACs was computed. The average proportion of non-zero RAC-RMFs based on shoes contributing a RAC 2.8 mm or longer across all three datasets was 8.4%.

Table 10: Comparison of RAC counts and the number of shoes with a non-zero RAC-RMF between the full database of test impressions (T) [6], the test impression subset for the blood dataset (TSB) [9], and the test impression subset for the dust dataset (TSD) at $t = 1.0$. For the subsample of RACs ≥ 2.8 mm, count is reported as a fraction of the total, followed by the percentage in parentheses.

Variable	**T**	**TSB**	**TSD**
Total Shoes in Dataset	1,300	162	162
Total RACs	80,668	18,887	19,169
Impressions with 1^+ RACs	1,300	122	146
RAC Subsample	—	759	1,513
Shoes with 1^+ RACs ≥ 2.8 mm	1,131	89	98
RACs ≥ 2.8 mm	19,069/80,668 (24%)	280/759 (37%)	425/1,513 (28%)
Shoes with Non-Zero RAC-RMF for RACs ≥ 2.8 mm	102/1,131 (9.0%)	7/89 (7.9%)	8/98 (8.2%)

Proportions tests [14] were used to compare the RAC counts and non-zero RAC-RMF values for the two test impression subsets to the full database, as shown in Table 11. It must be acknowledged that the assumption of the proportions test requiring random sampling was violated, but other assumptions were met. The results of these tests indicated that there were significantly higher proportions of RACs in the test impression subsets relative to the full database of test impressions with a length greater than or equal to 2.8 mm. This was expected since RAC subsampling was driven by features that could be detected above the signal-to-noise limit imparted by the medium and substrate in the associated dust or blood impressions. However, these tests failed to detect a significant difference in

the proportions of shoes with a non-zero RAC-RMF between the three datasets, indicating that the proportions were equivalent. This provides moderate support for the hypothesis that the number of RACs with a minimum length of 2.8 mm, when evaluated using the similarity metric, human observers, and mathematical model used in these studies, result in a relatively consistent proportion of non-zero RAC-RMFs. Strong support is not asserted since only three sample sets were evaluated and the datasets are clearly dependent on each other and the parameters/metrics of this investigation, but one could predict other subsamples of the WVU database to result in non-zero RAC-RMFs that follow the same trend (# Shoes with Non-Zero RAC-RMF = # Shoes with 1^+ RACs $\geq$ 2.8 mm $\times$ 8.4%). The natural next step would be to determine if this proportion (or a scaled multiple of a constant proportion) exists in other datasets analyzed using different metrics. If so, this would lend support for future predictions of the expected number of shoes with a non-zero RAC-RMF based on the number of shoes in a population contributing at least one RAC of the specified size.

Table 11: Results of proportions tests for the number of RACs with length greater than or equal to 2.8 mm and the number of shoes with a non-zero RAC-RMF between the full database of test impressions (T) [6], the test impression subset for the blood dataset (TSB) [9], and the test impression subset for the dust dataset (TSD) at $t = 1.0$.

		Count	**Proportion**	***p*-value**	**Significant?**	**Conclusion**
RACs $\geq$ 2.8 mm	T	19,069	0.24	< 0.0001	Yes	TSB > T
	TSB	280	0.37			
	T	19,069	0.24	< 0.0001	Yes	TSD > T
	TSD	425	0.28			
Shoes with Non-Zero RAC-RMF for RACs $\geq$ 2.8 mm	T	102	0.090	0.3557	No	TSB $\approx$ T
	TSB	7	0.079			
	T	102	0.090	0.3897	No	TSD $\approx$ T
	TSD	8	0.082			

3.7.2. Impressions in Simulated Crime Scene Datasets

By controlling for difference in RAC count with the test impression subset, further insight was gained regarding the influence of impression type and quality on RAC-RMF. Comparisons of the blood [9] and dust impressions to their respective test impression subsets are shown in Table 12. There was a 66% increase in the percentage of indistinguishable pairs in the blood impression dataset relative to the corresponding test impression subset, while there was a 42% decrease in the percentage of indistinguishable pairs in the dust impression dataset relative to its test impression subset. As a result, blood impressions had a 20% and 45% increase in the maximum number of shoes sharing at least one indistinguishable RAC pair with another shoe and the percentage of shoes with a non-zero RAC-RMF relative to TSB, respectively, while dust impressions had a 20% and 31% decrease in these variables relative to TSD. In other words, the RACs in blood impressions were more similar than their known mates to non-mated RACs from test impressions, while the opposite was true for the RACs from dust impressions. If this observation is generalized further, one could hypothesize that this was the result of a liquid medium negating RAC detail/edges, and therefore

limiting or lowering the weight of evidence of a single RAC identified in a questioned impression deposited in blood. In other words, liquid media may increase the frequency of encountering indistinguishable pairs in non-mated test impressions, and therefore result in a lower weight of evidence and/or fewer source association opinions when encountered in actual casework. Naturally, to be valid, this hypothesis must be further tested.

Table 12: Comparison of the results between blood [9] and dust simulated crime scene impressions (SCS) and their respective test impression subsets (TS) at $t \geq 0.5$. The percentage change was calculated with reference to the associated test impression subset using (SCS – TS)/TS × 100%, where a negative value indicates a decrease.

Variable	Impression Type	Blood	Dust
Indistinguishable Pairs	TS	0.06%	0.11%
	SCS	0.11%	0.07%
	% Change	66%	-42%
Maximum RAC-RMF (out of 1,299)	TS	5	10
	SCS	6	8
	% Change	20%	-20%
Shoes with Non-Zero RAC-RMF for Any Size RAC	TS	24%	47%
	SCS	34%	32%
	% Change	45%	-31%

3.7.3. RAC Size

Neither Part I [9] nor Part II of this investigation thoroughly investigated all the factors that might contribute to the degree of similarity between non-mated RACs. Attributes such as size, clarity, geometric complexity, and perhaps even category (linear, compact, variable) could be contributing factors, not to mention the strengths and limitations of the comparison metric (%A). Of all possible factors, only the size of the longest axis of an identified RAC was considered. Table 13 shows the results from four different proportions tests [14] performed between the number of RACs greater than or equal to 2.8 mm in length in the test impression subsets and simulated crime scene impressions for the blood [9] and dust datasets. Again, it must be acknowledged that the assumption requiring random sampling was violated, but other assumptions were met. The first test compared the two test impression subsets and concluded that there were significantly more RACs with length greater than or equal to 2.8 mm available in the blood test impression subset than in the dust test impression subset. The second test compared the proportion of RACs from the blood impressions and the corresponding test impression subset, and concluded that there were significantly more RACs longer than or equal to 2.8 mm available in the test impressions than the number of RACs of that size identified in the blood impressions. The third test performed a similar evaluation for dust impressions, which failed to detect a significant difference between impression types. The final test compared the two simulated crime scene impression datasets and concluded that there were significantly more RACs with length greater than or equal to 2.8 mm identified in the dust impressions than the blood impressions. In combination, these tests support two observations. First, the shoes selected to generate

the questioned blood impressions happened (by chance) to have fewer RACs overall (18,887 for blood versus 19,169 for dust), but a greater proportion that were "larger" (as defined by a major axis greater than or equal to 2.8 mm). Thus, there was a reasonable proportion of these larger RACs available for transfer when creating the blood impressions, but fewer were able to be detected, and/or if they transferred and were detected, they were eroded in size. Second, and conversely, of the proportion of the same "sized" RACs available for transfer from shoes selected to create the dust impressions, a relatively equal number were detected. This means that these features either transferred and were detected, and/or smaller features transferred but were dilated in size. These trends can also be seen in Table 2 by comparing the proportion of RACs of each impression type with a longest axis within a specified range. If this observation is extrapolated, barring confounding factors, it suggests that a liquid medium erodes RAC length, while a particulate medium either has no impact, or dilates RAC length, which is a possible explanation for the difference in the proportions of shoes with non-zero RAC-RMFs for the simulated crime scene impressions relative to the test impression subsets reported in Table 12.

Table 13: Results of proportions tests for the number of RACs with length greater than or equal to 2.8 mm in the test impression subsets (TS) and simulated crime scene impressions (SCS) for the blood (B) [9] and dust (D) datasets.

	Count	**Proportion**	***p*-value**	**Significant?**	**Conclusion**
TSB	280	0.37	< 0.0001	Yes	TSB > TSD
TSD	425	0.28			
TSB	280	0.37	< 0.0001	Yes	TSB > SCSB
SCSB	131	0.17			
TSD	425	0.28	0.1539	No	TSD ≈ SCSD
SCSD	400	0.26			
SCSB	131	0.17	< 0.0001	Yes	SCSD > SCSB
SCSD	400	0.26			

4. Discussion

In Part I of this study [9], blood was used to prepare a simulated crime scene footwear impression dataset to serve as a lower-quality complement to the high-quality test impressions analyzed in [6]. As a continuation, Part II repeated this analysis using a second simulated crime scene impression dataset composed of impressions deposited in dust. The motivation behind this comparison was to determine how, or if, findings vary across different media, substrates, and collection techniques. While Part I provided comparisons between the results of test and blood impressions, the following compares test versus dust impressions, as well as dust versus blood impressions and their corresponding test impression subsets.

For this study, 55 dust impressions were created on paper, 55 were created on tile and lifted with gelatin, and 55 were created on tile and lifted with Mylar film and an electrostatic lifter, totaling 165 impressions (reduced to 162 following registration to the corresponding

test impressions). While 146 (90%) of the shoes in the dataset had at least one RAC, over 92% of RACs did not transfer from the outsole to the corresponding dust impression. For comparison, 122 (75%) of the 162 shoes in the blood impression dataset had at least one RAC and over 95% of RACs did not transfer from the outsole to the corresponding blood impression [9]. Since the dust and blood impressions were generated from different outsoles, and varied in totality/partiality, clarity, and the removal/persistence of background interference in the form of a patterned-tile, it is possible to discuss an anticipated rate for features to transfer across a wide latitude of questioned impressions expected in casework. When the results from Part I [9] and Part II are combined with those reported in [15], which were based on impressions created in shoe polish, an approximate range between 5% and 15% of RACs present on outsoles are likely to transfer to questioned impressions of the quality and clarity analyzed in these studies. This demonstrates the highly probative value associated with forensic footwear analyses. In other words, the large percentage of shoes (83%) with at least one RAC that transferred and was detected in a simulated questioned impression, and the average count per impression type reported in Table 1 (ranging from 5 (blood) to 9 (all dust) with 95% confidence intervals between 2 and 17) is a strong argument for the collection of this type of evidence at crime scenes. In addition, more than one shared indistinguishable RAC pair was never identified between unrelated outsoles for the dust or blood [9] impression datasets. Thus, undetermined distinguishable RACs (m) are likely to exist in these questioned impressions, which would be factored into any estimate of impression-wide RMF, and therefore decrease the chance of an erroneous source association for any impressions with one or more RACs.

When comparing the size of all RACs in the 162 test impressions to those that reproduced in a dust impression, it was observed that larger RACs were more likely to transfer and be detected in a dust impression, similar to the blood impressions in Part I [9]. Comparison of the $\text{RACs}_{(\text{TSD})}$ in the test impression subset to their known mates in the dust impressions showed that RACs were larger in the dust impression 57% of the time (Fig. 7). However, this trend was most prevalent in the impressions lifted with gelatin, while the impressions lifted with Mylar film and on paper had relatively equal proportions of RACs that were larger versus those that were smaller in the dust impressions. One possible explanation for this result is that the adhesive and flexible nature of the gelatin lifter caused a slight increase in RAC size. More specifically, the gelatin lifter was pulled from the tile after collecting the impression, which could have led to slight stretching, despite avoiding the use of a fingerprint roller during collection and letting the lifter rest prior to photographing. However, 33% of the RACs present in the impressions lifted with gelatin were smaller than their test impression mate, so this observation could just as likely be attributed to variation in dust concentration/opacity to define the edges of a RAC. The overall 57% rate of increase in RAC size in dust impressions contrasted with the results of the blood impressions, where RACs were smaller than their test impression mate 66% of the time [9]. A possible explanation for this is that RACs deposited in liquid medium tend to appear eroded in size, while those deposited in particulate remain stable and/or appear dilated. While the largest size decrease observed was 4.9 mm^2 and the largest increase was 10.0 mm^2 (both for RACs in dust impressions), 95% of all RAC size differences were between –1.5 mm^2 and +1.6 mm^2 (please

see supplemental material section S9 and Fig. S9 for additional information). In addition, 80% of RACs in dust impressions and 76% of RACs in blood impressions maintained their original shape classification from the corresponding test impression. Thus, the collective results provide empirical evidence for examiners to claim that the majority of RACs do not change in size beyond ± 1.6 mm^2 when compared between test and questioned impressions, the rate of decrease in size is slightly elevated for blood impressions, and no more than about 25% of RACs will exhibit a geometric change substantial enough to shift classification between the categories of linear, compact, and variable, as assessed in this study.

RACs in blood impressions were found to exhibit higher percent area overlap values, higher probabilities of indistinguishability, more indistinguishable pairs, a higher maximum RAC-RMF value, and more non-zero RAC-RMF values when compared to non-mated $\text{RACs}_{(\text{T})}$ than observed in the corresponding test impression subset [9]. For the dust impressions, there was no noticeable shift in percent area overlap or probability of indistinguishability values for non-mated comparisons relative to the full database of test impressions [6] or its test impression subset (Fig. 9), and the dust impression dataset had less indistinguishable pairs, a lower maximum RAC-RMF, and less non-zero RAC-RMF values relative to its test impression subset, as outlined in Table 12. In other words, $\text{RACs}_{(\text{B})}$ were more similar to non-mated $\text{RACs}_{(\text{T})}$ than $\text{RACs}_{(\text{TSB})}$ were to non-mated $\text{RACs}_{(\text{T})}$, while $\text{RACs}_{(\text{D})}$ were less similar to non-mated $\text{RACs}_{(\text{T})}$ than $\text{RACs}_{(\text{TSD})}$ were to non-mated $\text{RACs}_{(\text{T})}$. This difference is hypothesized to be a function of relative RAC size, as evaluated in Table 13. More specifically, the $\text{RACs}_{(\text{B})}$ in blood impressions were smaller than their test impression mates and the $\text{RACs}_{(\text{D})}$ in dust impressions, making them more similar to binary digitized images of non-mated $\text{RACs}_{(\text{T})}$. If extrapolated to be representative of liquid versus particulate media, a possible forensic consequence of this observation is that liquid media may increase the frequency of encountering indistinguishable pairs in non-mated test impressions, and therefore result in a lower weight of evidence and/or fewer source association opinions when these types of questioned impressions are encountered in casework.

4.1. Limitations

The limitations of this study are similar to those outlined in detail in previous studies [6, 9]. To summarize, there are three limitations that must be considered when analyzing the high-quality test impressions in the WVU footwear database [7] and any other impressions processed using the same methodology. These include mixed makes and models of shoes, an outsole normalization procedure to re-map RACs to a standard shoe size and shape, and the spatial cells on the standardized outsole that were used to determine positional co-occurrence. In Part I of this investigation [9], five additional limitations were introduced when analyzing the simulated crime scene impressions made in blood, which also persist in this study. First, percent area overlap was again used to evaluate non-mated RACs for ease of comparison between datasets, despite previous research that showed that Hausdorff distance performed better when comparing RACs from simulated crime scene impressions to RACs from test impressions [8]. Second, pseudo-accidentals were prevented due to the modified sequential unmasking approach utilized during marking, and therefore were not included in RAC-RMF estimates. Third, this study focused on non-mated RAC pairs to

estimate RAC-RMF, with no evaluation of mated pairs to inform a likelihood ratio. Fourth, RAC-RMF estimates were based only on the occurrence of indistinguishable pairs and did not consider the number of distinguishable RACs also present, which would be required to obtain an impression-wide RMF estimate. Lastly, the overall reliability of RACs was not evaluated in detail for this database, and only the proxy of RAC length was considered. Within the context of a footwear reliability study [10], over 26% of the RACs in the dust impression dataset and 17% of the RACs in the blood impression dataset were longer than the average length (2.8 mm) of RACs deemed reliable by examiners. Using 2.8 mm as a threshold for reliable RACs on which to base opinions of source association/disassociation, 3 shoes out of 98 containing at least one reliable RAC (3.1%) presented non-zero RAC-RMFs for the dust impressions, while 3 shoes out of 89 containing at least one reliable RAC (3.4%) had non-zero RAC-RMFs for the blood impressions. Despite the similar values between datasets, there were twice as many RACs available in the dust impressions, indicating a decreased degree of similarity between $\text{RACs}_{(\text{D})}$ and $\text{RACs}_{(\text{T})}$ overall.

4.2. Future Considerations

In combination, analysis of blood [9] and dust impressions provided a range of estimates regarding RAC transfer and RAC-RMF that may be anticipated based on the quality and type of a questioned impression. While there are several possible avenues to continue the investigation of RAC-RMF in footwear, three areas of study should be prioritized. First, in order to achieve an impression-wide RMF estimate that could be useful in a casework scenario, the distinguishable RACs m present on unrelated outsoles must be considered. This would require an understanding of how n and m vary with the size and/or geometric complexity of the RACs present. Second, an investigation into the variation anticipated in known mate replicate crime scene impressions is needed, which could be used in combination with RMF estimates to inform a likelihood ratio. Lastly, these studies should be repeated using examiners to determine how well the estimates presented align with examiner evaluations of RAC similarity. The RAC comparisons performed herein were assessed by researchers, and based on features after they were extracted from impressions (please see supplemental material section S7 and Figs. S4 and S5 for examples). As a result, feature similarity was evaluated in the absence of impression context (*i.e.*, size of RAC with respect to a tread element, edge clarity and contrast with respect to background interference, etc.). If examiners rely on cues present in impressions to determine similarity when evaluating a pairwise comparison of RACs shown side-by-side, then including the background of the impression surrounding the RAC could influence these estimates — and quantifying this influence would be an important variable not yet considered.

Acknowledgments

The database used in this investigation was originally supported by Award No. 2013-DN-BX-K043, awarded by the National Institute of Justice (NIJ), Office of Justice Program, U.S. Department of Justice. The simulated crime scene dataset was supported by the Center for Statistics and Applications in Forensic Evidence (CSAFE), through Cooperative Agreement No. 70NANB20H019 between NIST and Iowa State University, which includes activities

carried out at West Virginia University. Images can be found in the CSAFE repository and are cited to Smale, A., Speir, J., West Virginia University 2022 High Quality and Simulated Crime Scene Dataset, Release #1; Release Date March 2022. In addition to the NIJ and CSAFE, the authors would like to thank Claire Dolton for registering the 165 dust impressions analyzed in this investigation. Lastly, the opinions, findings, conclusions, and recommendations expressed in this manuscript are those of the authors and do not necessarily reflect those of the Department of Justice, the Center for Statistics and Applications in Forensic Evidence, and/or the National Institute of Standards and Technology.

References

[1] M. J. Cassidy, Footwear Identification, Public Relations Branch of the Royal Canadian Mounted Police, Ottawa, Ontario, 1980.

[2] T. W. Adair, J. LeMay, A. McDonald, R. Shaw, R. Tewes, The Mount Bierstadt study: An experiment in unique damage formation in footwear, Journal of Forensic Identification 57 (2) (2007) 199–205.

[3] C. Hamburg, R. Banks, Evaluation of the random nature of acquired marks on footwear outsoles, Impression and Pattern Evidence Symposium; Clearwater, FL (2010).

[4] H. D. Wilson, Comparison of the individual characteristics in the outsoles of thirty-nine pairs of Adidas Supernova Classic shoes, Journal of Forensic Identification 62 (3) (2012) 194–203.

[5] M. Marvin, A look at close non-match footwear examinations, International Association for Identification (IAI) Centennial Conference; Cincinnati, OH (2015).

[6] A. N. Smale, J. A. Speir, Estimate of the random match frequency of acquired characteristics in a forensic footwear database, Science & Justice 63 (3) (2023) 427–437. doi:https://doi.org/10.1016/j.scijus.2023.04.007.

[7] J. A. Speir, N. Richetelli, M. Fagert, M. Hite, W. J. Bodziak, Quantifying randomly acquired characteristics on outsoles in terms of shape and position, Forensic Science International 266 (2016) 399–411. doi:https://doi.org/10.1016/j.forsciint.2016.06.012.

[8] N. Richetelli, W. J. Bodziak, J. A. Speir, Empirically observed and predicted estimates of chance association: Estimating the chance association of randomly acquired characteristics in footwear comparisons, Forensic Science International 302 (2019) 1–14. doi:https://doi.org/10.1016/j.forsciint.2019.05.049.

[9] A. N. Smale, J. A. Speir, Estimate of the random match frequency of acquired characteristics in footwear: Part I — Simulated crime scene impressions in blood [Manuscript submitted for publication] (2023).

[10] N. Richetelli, L. Hammer, J. A. Speir, Forensic footwear reliability: Part II—Range of conclusions, accuracy, and consensus, Journal of Forensic Sciences 65 (6) (2020) 1871–1882. doi:https://doi.org/10.1111/1556-4029.14551.

[11] E. T. Lin, T. DeBat, J. A. Speir, A simulated crime scene footwear impression database for teaching and research purposes, Journal of Forensic Sciences 67 (2022) 726–734. doi:https://doi.org/10.1111/1556-4029.14933.

[12] F. Garwood, Fiducial limits for the Poisson distribution, Biometrika 28 (3/4) (1936) 437–442. doi:https://doi.org/10.2307/2333958.

[13] W. J. Bodziak, Footwear Impression Evidence: Detection, Recovery, and Examination, 2nd Edition, CRC Press, Boca Raton, FL, 2000.

[14] D. C. Montgomery, G. C. Runger, Applied Statistics and Probability for Engineers, 6th Edition, John Wiley & Sons, Hoboken, NJ, 2013.

[15] N. Richetelli, M. Nobel, W. J. Bodziak, J. A. Speir, Quantitative assessment of similarity between randomly acquired characteristics on high quality exemplars and crime scene impressions via analysis of feature size and shape, Forensic Science International 270 (2017) 211–222. doi:https://doi.org/10.1016/j.forsciint.2016.10.008.

Estimate of the Random Match Frequency of Acquired Characteristics in Footwear: Part II — Simulated Crime Scene Impressions in Dust

Supplemental Material

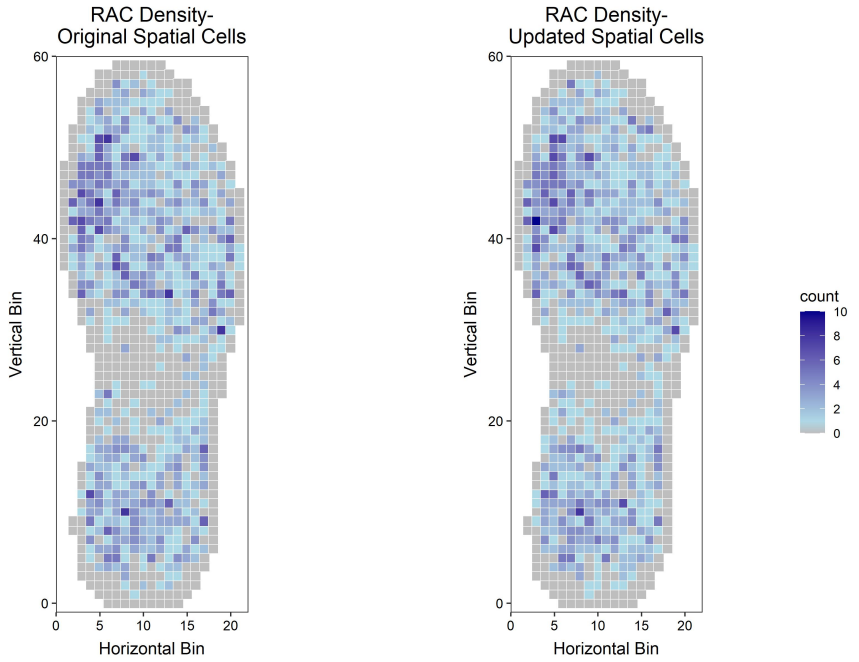
S1. Lifted Dust Impressions

Dust impressions were lifted with Mylar film and an electrostatic dust lifter according to standard procedure as described by Bodziak (2000) and Lin *et al.* (2022). Briefly, Mylar film was placed over the dust impression, and the electrostatic lifter was positioned so that it was touching both the Mylar film and a metal grounding plate. Once the voltage was applied, a fingerprint roller was used to smooth the Mylar film. After reducing the charge, the Mylar film was lifted from the tile beginning from one of the corners.

As there are many method variations possible when utilizing a gelatin lifter, further elaboration on technique is required. The collection of dust impressions using a gelatin lifter was not aided by the use of a fingerprint roller. Instead, dust impressions deposited on tile were lifted by laying a gelatin lifter on the impression and applying weight for a duration of 5 minutes. More specifically, the gelatin was held by its ends and without tension in order to create a U-shape. The center of the gelatin lifter (apex of the "U") was allowed to contact the tile first, before both ends were released, essentially unrolling the U-shaped gelatin lifter from its center toward its edges. Next, a full ream of 8.5" × 11" copy paper was placed on top of the gelatin lifter for no less than 5 minutes. The gelatin lifter was then pulled up from the tile beginning from one of the corners and allowed to rest for at least 5 minutes before being photographed.

S2. Spatial Distribution of RACs in the Dust Impressions

The 1,513 $\text{RACs}_{(\text{D})}$ identified in the dust impressions were extracted and localized to a spatial cell on the standardized outsole based on the RAC's centroid, as described in Speir *et al.* (2016). The $\text{RACs}_{(\text{D})}$ were distributed in 630 out of 987 cells with a maximum of 8 RACs per cell, as shown in Fig. S1(a). The resulting location of each $\text{RAC}_{(\text{D})}$ was compared to its known mate in the corresponding test impression, and it was determined that 257 $\text{RACs}_{(\text{D})}$ (17%) had shifted into a different cell in the dust impression due to either the RAC's change in shape resulting in a change in its centroid, and/or slight misalignments during registration. In order for each $\text{RAC}_{(\text{D})}$ from the dust impression to be compared to the same non-mated $\text{RACs}_{(\text{T})}$ as its mate from the test impression, each of these 257 $\text{RACs}_{(\text{D})}$ was re-assigned to the original cell of its known mate. As a result of this adjustment, 1,513 $\text{RACs}_{(\text{D})}$ were distributed in 645 out of 987 cells with a maximum of 10 RACs per cell, as shown in Fig. S1(b).



(a) (b)

Figure S1: Distribution of $\text{RACs}_{(\text{D})}$ from dust impressions in the original spatial cells based on each RAC's centroid (a) and the updated cells using the mated $\text{RAC}_{(\text{TSD})}$ from the test impression as ground truth (b).

S3. RAC Count by Impression Type

Fig. S2 plots the RAC count in the test impression versus the dust impression for all 162 shoes in the simulated crime scene dataset. Different colors are used to represent the three types of dust impressions.

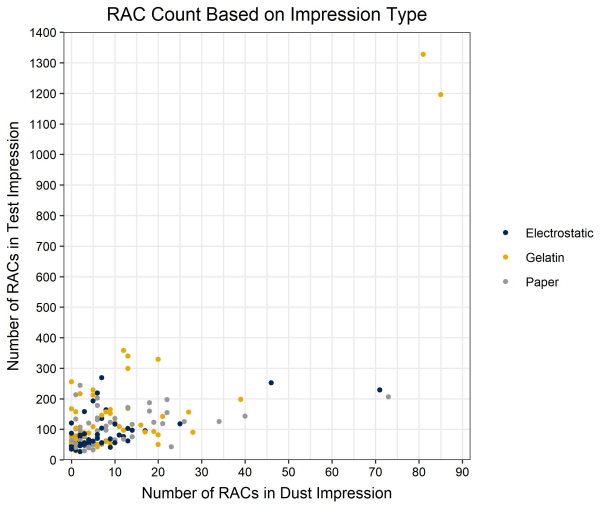


Figure S2: Number of RACs identified in the dust impression versus the number of RACs previously identified in the high-quality test impression for all 162 shoes in the simulated crime scene impression dataset.

S4. Known Mate RAC Areas

The area of each RAC in the dust impressions and its mate in the corresponding test impression was computed, and all 1,513 RACs are plotted in Fig. S3. Approximately 57% of RACs were larger in the dust than in the test impression, while 42% were smaller and 6 RACs (<0.5%) were the same size in both impressions.

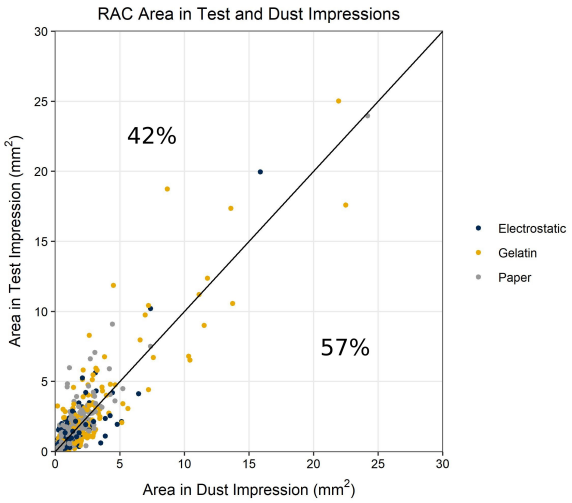


Figure S3: Relative area of known mated RACs in test and dust impressions. The diagonal line represents no change in RAC size between impression types.

S5. Quality Control

There were six impressions of each impression type (electrostatic, gelatin, and paper) which were blindly marked twice to assess researcher variation in marking. The RAC count information for these 18 impressions is shown in Table S1, where the letter in parentheses following each shoe number indicates the impression type. During post-processing the duplicate impressions were denoted as "Marking A" and "Marking B" where "A" reflects the first time an impression was marked and "B" represents the second. Using the fourth row of the table for shoe 265R (electrostatic) as an example, 12 RACs (none of which were biased) were identified during the first mark-up (A) and 16 RACs (2 of which were biased) were identified in the second attempt (B), for a count difference of 4 RACs and a percentage difference out of the maximum number found of 75% (or (12/16) × 100%). When comparing the RACs identified during each marking, 10 RACs were consistent in both markings, which is 71.4% of the total number of RACs found (or (10+10)/(12+16) × 100%).

Table S1: RAC counts from duplicate markings of the 18 impressions used as part of the quality control program. For RAC and bias counts, the first number corresponds to Marking A and the second corresponds to Marking B.

Shoe ID	RAC Count (A/B)	Bias Count (A/B)	Count Difference	RACs in Both	Percentage of Max. (%)	Percentage in Both (%)
034R (E)	3/3	0/0	0	3	100	100
057R (E)	2/1	0/0	1	1	50.0	66.7
165L (E)	1/1	0/0	0	1	100	100
265R (E)	12/16	0/2	4	10	75.0	71.4
582L (E)	0/2	0/0	2	0	0	0
632R (E)	3/4	0/0	1	2	75.0	57.1
073R (G)	13/7	1/0	6	2	53.8	20.0
082L (G)	0/0	0/0	0	0	—	—
085R (G)	3/7	0/0	4	3	42.9	60.0
093R (G)	81/111	5/4	30	74	73.0	77.1
404L (G)	2/1	0/0	1	1	50.0	66.7
411R (G)	15/28	0/0	13	13	53.6	60.5
018L (P)	5/5	0/0	0	5	100	100
049R (P)	14/10	1/3	4	7	71.4	58.3
054L (P)	6/9	0/1	3	6	66.7	80.0
107R (P)	22/21	0/0	1	20	95.5	93.0
189L (P)	2/2	0/0	0	2	100	100
490R (P)	22/26	0/3	4	20	84.6	83.3

S6. Uncertainty in Visual Comparisons

There were 3,765 visual comparisons performed between $\text{RACs}_{(\text{TSD})}$ and $\text{RACs}_{(\text{T})}$, and another 3,765 visual comparisons performed between $\text{RACs}_{(\text{D})}$ and $\text{RACs}_{(\text{T})}$ for this study. These comparisons were performed in triplicate by two researchers. During these repeated trials, a total of 138 RAC pairs for the test impression subset were deemed indistinguishable, while an additional 34 pairs were sometimes classified as indistinguishable and sometimes classified as exhibiting minor differences, and therefore were deemed uncertain. For the dust impression dataset, 72 pairs were consistently deemed indistinguishable and 25 pairs were uncertain. In order to report a worst-case scenario for this study, these uncertain pairs were included as indistinguishable pairs for all analyses (for a total of 172 and 97 pairs, respectively).

S7. Indistinguishable Pairs

A selection of the 44 RAC pairs previously deemed indistinguishably through visual comparison by Richetelli *et al.* (2019) using "Test Criteria" are shown in Fig. S4, where both RACs are from unrelated test impressions. For comparison, Fig. S5 shows a selection of additional indistinguishable pairs using the more lenient "CS Criteria" during visual comparison, where the black RAC is from a dust impression and the gray RAC is from an unrelated test impression.

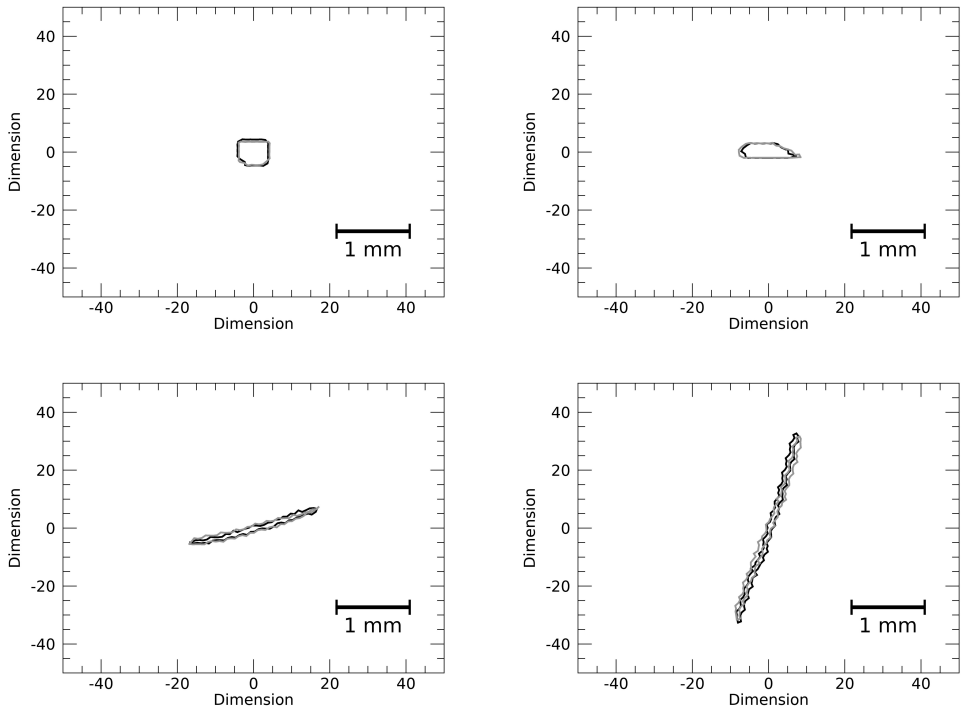


Figure S4: A sampling of non-mated $\text{RACs}_{(\text{T})}$ deemed indistinguishable using Test Criteria.

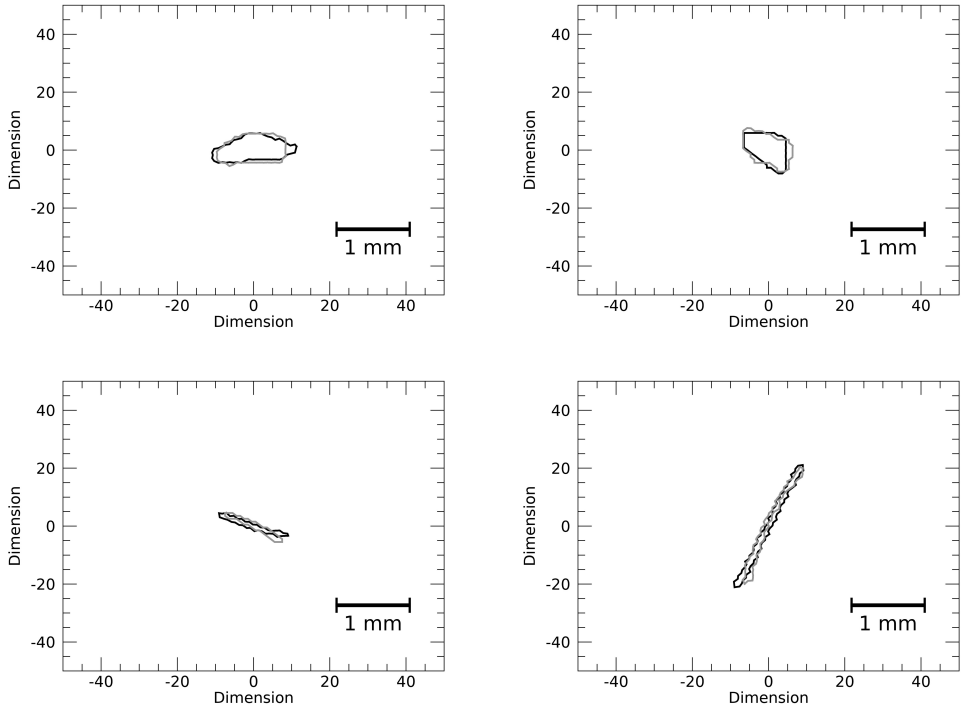
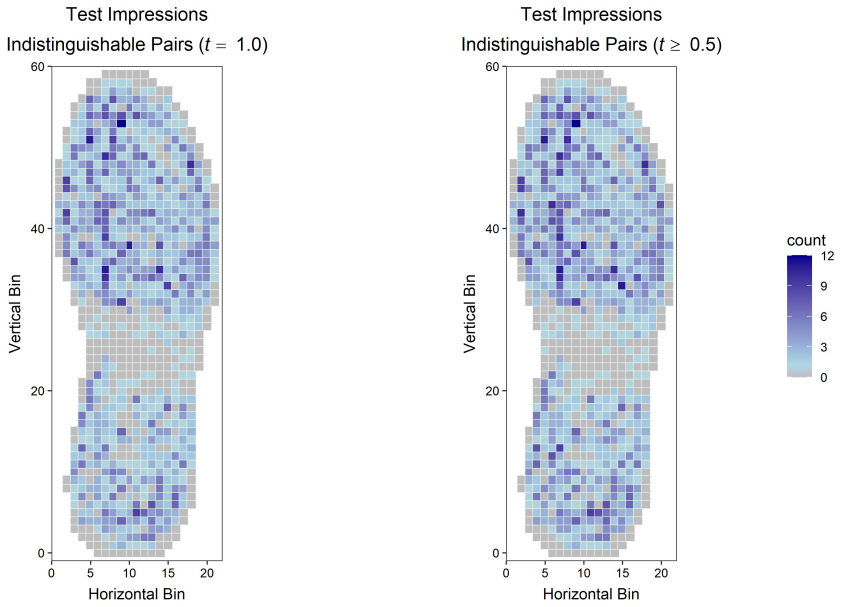


Figure S5: A sampling of $\text{RACs}_{(\text{D})}$ (black) and $\text{RACs}_{(\text{T})}$ (gray) deemed indistinguishable using CS Criteria.

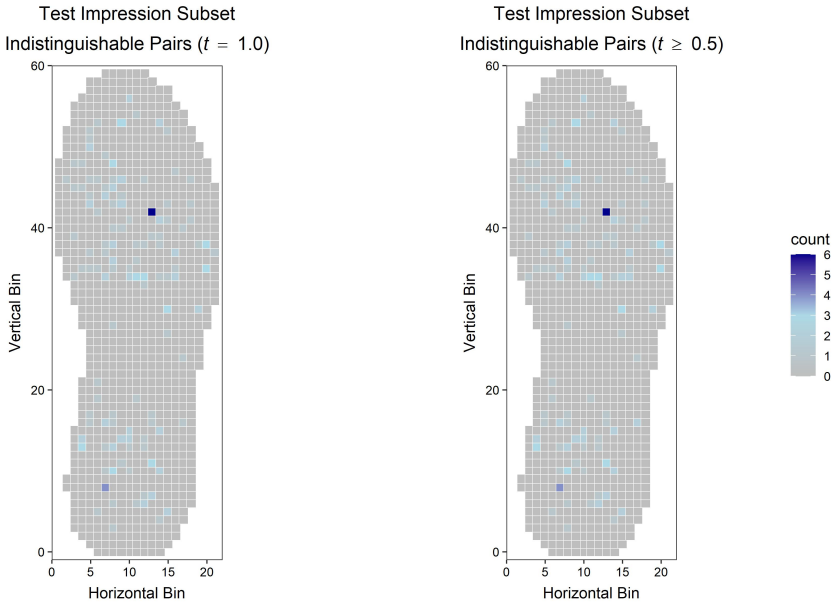
S8. Spatial Distribution of Indistinguishable Pairs

The spatial distribution of the indistinguishable RAC pairs across the standardized outsole is displayed for the full database of test impressions (Fig. S6) as discussed by Smale and Speir (2023), the test impression subset (Fig. S7), and the dust impressions (Fig. S8) for $t = 1.0$ and $t \geq 0.5$. For the full database of test impressions, there were 2,181 *certain* indistinguishable pairs at a threshold of $t = 1.0$ and 2,336 *possible* indistinguishable pairs at $t \geq 0.5$. There were 172 *certain* and 175 *possible* indistinguishable pairs for the test impression subset. Lastly, there were 97 *certain* and 101 *possible* indistinguishable pairs for the dust impressions.



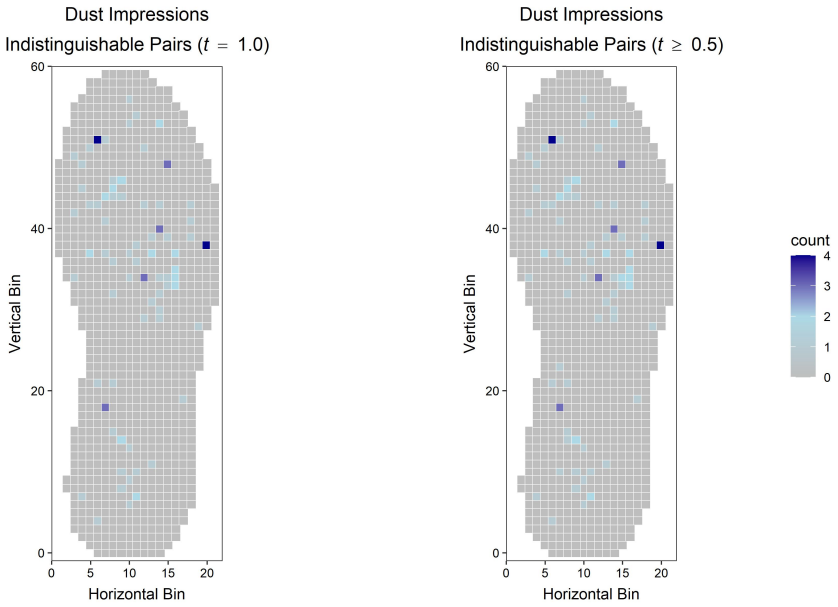
(a) (b)

Figure S6: Heatmaps showing the distribution of indistinguishable pairs across the outsole from the full database of high-quality test impressions for $t = 1.0$ (a) and $t \geq 0.5$ (b), reproduced from Smale and Speir (2023).



(a) (b)

Figure S7: Heatmaps showing the distribution of indistinguishable pairs across the outsole for the test impression subset for $t = 1.0$ (a) and $t \geq 0.5$ (b). Note that these pairs have been plotted using a different density scale than Fig. S6 and Fig. S8.



(a) (b)

Figure S8: Heatmaps showing the distribution of indistinguishable pairs across the outsole from dust impressions for $t = 1.0$ (a) and $t \geq 0.5$ (b). Note that these pairs have been plotted using a different density scale than Fig. S6 and Fig. S7.

S9. RAC Size Difference by Impression Type

Fig. S9 plots the size difference of known mated RACs in the test impression subset versus the simulated crime scene impression for blood and dust. RACs in blood impressions were often smaller than their test impression mates, which is observed in the plot by the shift toward positive size differences. Conversely, RACs from dust impressions are shifted toward negative values. For blood impressions, the largest decrease was 3.0 mm^2 and the largest increase was 8.8 mm^2. The largest decrease for dust impressions was 4.9 mm^2 and the largest increase was 10.0 mm^2. Despite these extremes, 95% of all RAC size differences were between –1.5 mm^2 and +1.6 mm^2.

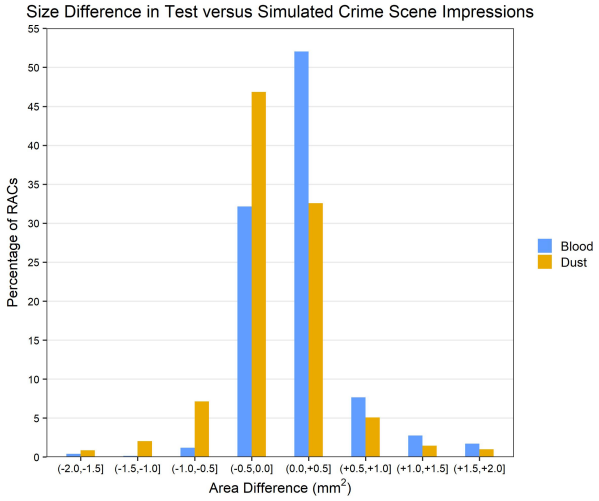


Figure S9: Size difference of known mated RACs in the test impression subset versus the simulated crime scene impression for blood and dust. While only differences between –2.0 mm^2 and +2.0 mm^2 are plotted, this includes 98% of the data. It must be acknowledged that some degree of bias in marking could be present in these observations, and if so, would likely contribute to the low size differences overall.

References

* W. J. Bodziak, Footwear Impression Evidence: Detection, Recovery, and Examination, 2nd Edition, CRC Press, Boca Raton, FL, 2000.
* E. T. Lin, T. DeBat, J. A. Speir, A simulated crime scene footwear impression database for teaching and research purposes, Journal of Forensic Sciences 67 (2022) 726–734. doi:https://doi.org/10.1111/1556-4029.14933.
* J. A. Speir, N. Richetelli, M. Fagert, M. Hite, W. J. Bodziak, Quantifying randomly acquired characteristics on outsoles in terms of shape and position, Forensic Science International 266 (2016) 399–411. doi:https://doi.org/10.1016/j.forsciint.2016.06.012.
* N. Richetelli, W. J. Bodziak, J. A. Speir, Empirically observed and predicted estimates of chance association: Estimating the chance association of randomly acquired characteristics in footwear comparisons, Forensic Science International 302 (2019) 1–14. doi:https://doi.org/10.1016/j.forsciint.2019.05.049.
* A. N. Smale, J. A. Speir, Estimate of the random match frequency of acquired characteristics in a forensic footwear database, Science & Justice 63 (3) (2023) 427–437. doi:https://doi.org/10.1016/j.scijus.2023.04.007.

5. Conclusions

This research had three main objectives. The first was to perform an evaluation of random match frequency of randomly acquired characteristics in a large database of test impressions. The second was to repeat the analysis using two smaller datasets of simulated crime scene impressions deposited in blood and dust on a variety of substrates. Upon completion of these first two goals, the third objective was to compare the RAC-RMF estimates associated with impressions of different quality and type. The results of this investigation provided valuable information for the forensic footwear community regarding the similarity of RACs that can occur on non-mated outsoles and the influence of impression medium and/or substrate on these findings.

5.1 High-Quality Test Impressions

High-quality test impressions (T) were previously created for each of the 1,300 shoes in the WVU footwear database [12], which provided a useful starting point to estimate RAC-RMF. Each shoe in the database was sequentially held out and compared to the remaining 1,299 outsoles to determine if unrelated shoes possessed similar $\text{RACs}_{(\text{T})}$ in the same locations, as determined by 5 mm $\times$ 5 mm spatial cells on the standardized outsole [12]. With over 80,000 $\text{RACs}_{(\text{T})}$ available for analysis, this resulted in nearly four million comparisons. In an attempt to determine indistinguishability with a reasonable level of efficiency, the comparisons were performed using a combination of two methods. In a previous study [13], 91,600 of these non-mated $\text{RAC}_{(\text{T})}$ pairs were visually-compared by human observers, and the results regarding indistinguishability were retained for the purposes of this research. For the remaining non-mated pairs, a mathematical model was used to predict the probability of indistinguishability based on the pair's percent area overlap similarity score [13].

While more than 99% of non-mated $\text{RACs}_{(\text{T})}$ were deemed distinguishable by these methods, indistinguishable $\text{RAC}_{(\text{T})}$ pairs were detected between unrelated outsoles. *Certain* indistinguishable pairs were those confirmed by visual comparison, while lower probability of indistinguishability thresholds were used to denote *plausible* and *possible* indistinguishable pairs as predicted by the mathematical model. These indistinguishable pairs were then used to estimate $\text{RAC}_{(\text{T})}$-RMF. It was concluded that nearly 70% of the shoes in the database had a $\text{RAC}_{(\text{T})}\text{-RMF}_{(m|n \geq 1)}$ greater than or equal to 1 out of 1,299. In other words, these shoes shared *at least one* indistinguishable RAC ($n \geq 1$), and an undetermined number of distinguishable RACs (m), with 1 or more of the 1,299 unrelated shoes in the database. In addition, the maximum $\text{RAC}_{(\text{T})}\text{-RMF}_{(m|n \geq 1)}$ value observed was 49 out of 1,299, meaning that a shoe shared at least one indistinguishable RAC with 49 unrelated shoes. Each pair of shoes with at least one shared indistinguishable $\text{RAC}_{(\text{T})}$ pair was further evaluated, and the maximum number of shared indistinguishable pairs for this database was $n = 3$ for *certain* and *plausible* indistinguishable pairs and $n = 5$ for *possible* indistinguishable pairs.

This analysis demonstrated that RAC geometries do repeat with positional similarity on unrelated outsoles, contrary to the majority of past work [7–11]. Although the impressions analyzed

were of much higher quality than those encountered in casework by subject matter experts, these results provided a point of reference for RAC-RMF estimates, which were previously believed to be zero (or near-zero) based on past empirical studies and theoretical models. These results prompted further investigation using simulated crime scene impressions to determine the effect of impression quality on RAC transfer and RAC-RMF estimates.

5.2 Simulated Crime Scene Impressions

Two simulated crime scene impression datasets were created for the purpose of estimating RAC-RMF in lower quality impressions. The first consisted of more than 160 simulated crime scene impressions deposited in blood (B) on different tile substrates and enhanced with leucocrystal violet. Approximately 75% of impressions had at least one RAC, with an average of five RACs per shoe and a total of 759 $\text{RACs}_{(\text{B})}$ identified. The second dataset included more than 160 simulated crime scene impressions deposited in dust (D) on paper and tile, with the latter lifted using either gelatin or Mylar film and an electrostatic lifter. Combining the three types of dust impressions, 90% had at least one RAC, with an average of nine RACs per shoe and a total of 1,513 $\text{RACs}_{(\text{D})}$. When considering the results of simulated crime scene impressions made in blood, dust, and shoe polish [14], a possible rate of RAC transfer between 5% and 15% can be anticipated. The known mate of each RAC from the blood and dust impressions was identified in the corresponding test impression from the same shoe to form a test impression subset (TS) for each dataset (referred to as TSB for the blood dataset and TSD for the dust dataset). These subsets allowed for investigation into the effect of dataset size relative to the full database of test impressions as well as the influence of impression quality relative to the simulated crime scene impressions.

Each RAC in the simulated crime scene impressions was compared to $\text{RACs}_{(\text{T})}$ with positional similarity in non-mated test impressions from 1,299 outsoles. Over 77,000 non-mated RAC comparisons were performed between blood impressions and non-mated test impressions. With fewer RACs and 50× less comparisons relative to the full database of test impressions, the number of indistinguishable pairs and the overall magnitude of RAC-RMF values decreased. When compared to their known mate from the test impression subset, $\text{RACs}_{(\text{B})}$ in blood impressions were found to be overall smaller in size. As a result, a 66% increase in the number of indistinguishable pairs was observed when comparing $\text{RACs}_{(\text{B})}$ in blood impressions to $\text{RACs}_{(\text{T})}$ in unrelated test impressions than when comparing $\text{RACs}_{(\text{TSB})}$ to the same non-mated $\text{RACs}_{(\text{T})}$. This translated into a higher proportion of shoes with a non-zero $\text{RAC}_{(\text{B})}$-RMF for the blood impressions (34%) than the test impression subset (24%) when including RACs of any size. Based on the average size of RACs deemed reliable by examiners in a recent footwear reliability study [16], a size threshold of 2.8 mm was implemented. Indistinguishable RAC pairs larger than this threshold existed in the blood impression dataset, and non-zero RAC-RMFs were observed at a rate of 3.4% when restricted to shoes with at least one RAC with a length greater than or equal to 2.8 mm, and 2.5% for shoes with at least one RAC of any length.

For the dust impressions, over 154,000 non-mated RAC comparisons were performed, which was a 25× decrease from the number of comparisons in the full database of test impressions. In contrast to the blood impressions, the $\text{RACs}_{(\text{D})}$ in dust impressions were often similar in size or larger than their known mates, resulting in a 42% decrease in the number of indistinguishable pairs relative to the corresponding test impression subset. Thus, a lower proportion of shoes with a non-zero RAC-RMF was observed for the dust impressions (32%) compared to the test impression subset (47%) when including RACs of any size. Using the size threshold of 2.8 mm, non-zero $\text{RAC}_{(\text{D})}$-RMFs were encountered at a rate of 3.1% for this dataset when restricted to shoes with at least one RAC with a length greater than or equal to 2.8 mm, and 2.1% for shoes with at least

one RAC of any length. While these values are similar to those observed for the blood impressions, the results highlight the decrease in overall similarity of non-mated RACs for dust impressions since there were twice as many RACs available. For both simulated crime scene datasets, no more than one shared indistinguishable RAC pair ($n = 1$) was ever observed between unrelated outsoles. Given that the average RAC count per impression ranged from 5 (blood) to 9 (dust) with 95% confidence intervals between 2 and 17, there are undetermined distinguishable RACs (m) in these questioned impressions which would likely decrease the chance of an erroneous source association for any impressions with one or more RACs.

While there are likely several confounding factors impacting the results obtained from test, blood, and dust impressions, three trends were observed to be associated with RAC-RMF. First, the proportions of shoes with a non-zero RAC-RMF considering only shoes with RACs with length greater than or equal to 2.8 mm were found to be equivalent across the full database of test impressions and the two test impression subsets, with an average of 8.4% of shoes. While this value should not be extrapolated to other datasets without additional study, it was hypothesized that this rate would likely be consistent across other subsamples of the WVU database. Second, the $\text{RACs}_{(\text{B})}$ from blood impressions analyzed in this study were more similar to non-mated $\text{RACs}_{(\text{T})}$ from test impressions relative to the corresponding test impression subset, while the $\text{RACs}_{(\text{D})}$ from dust impressions were less similar. This finding suggests a lower weight of evidence may be associated with $\text{RACs}_{(\text{B})}$ in blood impressions. Lastly, as a possible explanation for the increased similarity observed with blood impressions, RAC size was investigated. It was determined that the use of a liquid medium may erode RAC size in blood impressions, while a particulate medium maintains or possibly increases RAC size. As a result, RACs were overall smaller in blood impressions than dust impressions, making $\text{RACs}_{(\text{B})}$ more similar to non-mated $\text{RACs}_{(\text{T})}$ when comparing binary digitized images. Despite the influence of these factors on RAC size, the majority of RAC areas were not changed beyond ± 1.6 mm^2, and only about 25% of RACs exhibited a geometric change substantial enough to shift classification between the categories of linear, compact, and variable.

5.3 Future Considerations

While this research provided a starting point for understanding RAC-RMF, additional areas of study remain. First, in order to achieve an impression-wide RMF estimate that could be useful in a casework scenario, the distinguishable RACs m present on unrelated outsoles must be considered. This would require an understanding of how n and m vary with the size and/or geometric complexity of the RACs present. Second, an investigation into the variation anticipated in known mate replicate crime scene impressions is needed, which could be used in combination with RMF estimates to inform a likelihood ratio. Lastly, these studies should be repeated using examiners to determine how well the estimates presented align with examiner evaluations of RAC similarity. The RAC comparisons performed herein were assessed by researchers, and based on features after they were extracted from impressions. As a result, feature similarity was evaluated in the absence of impression context (*i.e.*, size of RAC with respect to a tread element, edge clarity and contrast with respect to background interference, etc.). If examiners rely on cues present in impressions to determine similarity when evaluating a pairwise comparison of RACs shown side-by-side, then including the background of the impression surrounding the RAC could influence these estimates — and quantifying this influence would be an important variable not yet considered.

Overall, this research demonstrated that RACs with positional and geometric similarity occur on unrelated outsoles at a more frequent rate than any estimates previously reported. However, the results presented are specific to these shoes and methods of analysis. As demonstrated with the simulated crime scene datasets, different experimental conditions can drastically affect RAC

transfer and RAC-RMF estimates. Therefore, continued work is required before rates can be fully understood and reported in forensic casework. Instead, the results should be used to calibrate footwear examiners when forming subjective opinions, particularly with regard to the fact that non-mated RACs can appear similar to each other, and that the medium and/or substrate of a questioned impression may influence the rate at which this occurs.

Bibliography

[1] W. J. Bodziak, Footwear Impression Evidence: Detection, Recovery, and Examination, 2nd Edition, CRC Press, Boca Raton, FL, 2000.

[2] SWGTREAD, Range of Conclusions Standard for Footwear and Tire Impression Examinations, https://www.nist.gov/system/files/documents/2016/10/26/swgtread_10_range_of_conclusions_standard_for_footwear_and_tire_impression_examinations_201303.pdf, 2013 (accessed June 2023).

[3] National Research Council: Committee on DNA Forensic Science, The Evaluation of Forensic DNA Evidence, National Academy Press, Washington, D.C., 1996.

[4] A. S. Fawcett, The role of the footmark examiner, Journal of the Forensic Science Society 10 (4) (1970) 227–244. doi:https://doi.org/10.1016/S0015-7368(70)70613-0.

[5] R. S. Stone, Footwear examinations: Mathematical probabilities of theoretical individual characteristics, Journal of Forensic Identification 56 (4) (2006) 577–599.

[6] Y. Yekutieli, Y. Shor, S. Wiesner, T. Tsach, Expert assisting computerized system for evaluating the degree of certainty in 2D shoeprints, Technical Report. National Institute of Justice, U.S. Department of Justice, Office of Justice Programs (2012).

[7] M. J. Cassidy, Footwear Identification, Public Relations Branch of the Royal Canadian Mounted Police, Ottawa, Ontario, 1980.

[8] T. W. Adair, J. LeMay, A. McDonald, R. Shaw, R. Tewes, The Mount Bierstadt study: An experiment in unique damage formation in footwear, Journal of Forensic Identification 57 (2) (2007) 199–205.

[9] C. Hamburg, R. Banks, Evaluation of the random nature of acquired marks on footwear outsoles, Impression and Pattern Evidence Symposium; Clearwater, FL (2010).

[10] H. D. Wilson, Comparison of the individual characteristics in the outsoles of thirty-nine pairs of Adidas Supernova Classic shoes, Journal of Forensic Identification 62 (3) (2012) 194–203.

[11] M. Marvin, A look at close non-match footwear examinations, International Association for Identification (IAI) Centennial Conference; Cincinnati, OH (2015).

[12] J. A. Speir, N. Richetelli, M. Fagert, M. Hite, W. J. Bodziak, Quantifying randomly acquired characteristics on outsoles in terms of shape and position, Forensic Science International 266 (2016) 399–411. doi:https://doi.org/10.1016/j.forsciint.2016.06.012.

[13] N. Richetelli, W. J. Bodziak, J. A. Speir, Empirically observed and predicted estimates of chance association: Estimating the chance association of randomly acquired characteristics

in footwear comparisons, Forensic Science International 302 (2019) 1–14. doi:https://doi.org/10.1016/j.forsciint.2019.05.049.

[14] N. Richetelli, M. Nobel, W. J. Bodziak, J. A. Speir, Quantitative assessment of similarity between randomly acquired characteristics on high quality exemplars and crime scene impressions via analysis of feature size and shape, Forensic Science International 270 (2017) 211–222. doi:https://doi.org/10.1016/j.forsciint.2016.10.008.

[15] S. Wiesner, Y. Shor, T. Tsach, N. Kaplan-Damary, Y. Yekutieli, Dataset of digitized RACs and their rarity score analysis for strengthening shoeprint evidence, Journal of Forensic Sciences 65 (3) (2020) 762–774. doi:https://doi.org/10.1111/1556-4029.14239.

[16] N. Richetelli, L. Hammer, J. A. Speir, Forensic footwear reliability: Part II—Range of conclusions, accuracy, and consensus, Journal of Forensic Sciences 65 (6) (2020) 1871–1882. doi:https://doi.org/10.1111/1556-4029.14551.