

Robert B. Kennedy,<sup>1</sup> Sanping Chen,<sup>2</sup> Ph.D.; Irwin S. Pressman,<sup>3</sup> Ph.D.;  
A. Brian Yamashita,<sup>1</sup> Ph.D.; and Ari E. Pressman,<sup>4</sup> M.D., FRCSC

## A Large-Scale Statistical Analysis of Barefoot Impressions\*

**ABSTRACT:** In an earlier paper, outlines of footprints of persons walking normally were studied to determine whether different people make verifiably distinct footprints. Our basic null hypothesis is: given a footprint outline trace made by Subject A (Alice), then Subject B (Bob), a distinct person, cannot produce a footprint outline trace indistinguishable from that of Alice. We showed in the previous work that the probability of a chance match is less than  $10^{-8}$ . In this paper we report two new advances in our research. First, we establish a rigorous mathematical framework for calculating worstcase and average chance-match probabilities. Second, we repeat the previous experiment to substantiate the earlier results, but with an expanded population sample size and a more representative and significantly bigger repeated sample. These improvements and a new automated tracing procedure for extracting all numerical measures lead to a sharpened accuracy with average chance match probabilities of  $7.88 \times 10^{-10}$  for a general population. In other words, the odds of a chance match are one in 1.27 billion.

**KEYWORDS:** forensic science, barefoot impressions, footprint, principal component analysis, chance match probability, statistical independence

Forensic barefoot morphology comparison examines the shapes of barefoot impressions in order to determine whether a suspect should be included or excluded as the person who made a footprint linked to a crime scene (1–5). In some cases, the perpetrator of a crime may have walked through the scene in bare or socked feet. In other cases, footwear impressions can be positively identified to a crime scene. When a suspect is arrested, his barefoot impressions can be compared to the bare or socked impressions at the crime scene, or to impressions left inside footwear that have been identified (6).

As with any individualizing technique, the question of primary importance concerns the fundamental uniqueness of the characteristic being compared. If ridge detail is visible in a barefoot impression at the crime scene, then the identification of a suspect can be made based on the principles used in fingerprint identification. However, when it is simply the shapes of the weight-bearing areas of a foot that are being considered, the underlying question of uniqueness or rarity must be established (5,7–10).

The notion that “everyone is different” arises when we look through a high school yearbook and appreciate that the human eye can readily distinguish one person from another. Even in the case of monozygotic twins, family members and close friends readily

discriminate subtle differences in appearance. Thus, morphological characteristics of an individual’s appearance can be used to identify them for social reasons, or for forensic applications.

The human skeleton is made up of 206 bones, 26 in each foot. There are 20 muscles in each foot. There are considerable variations in the component parts of the foot and ankle in terms of bone, muscle, ligaments, and tendons. These variations are reflected in the degree of contribution of each part to the function of the foot as a whole (11). The degree of contribution of each individual component is considered integral to the connected function of the foot during gait. Approximately 63 PSI peak pressure ( $43.3 \text{ N/cm}^2$ ) is placed on the bottom of an adult foot while walking normally (12). The progression of weight bearing across the normal foot during gait has been well documented (13). Similarly, the pressure distribution of weight through gait has been studied using footprints that were created using a barograph (14–16). Disease, injury or surgery can cause differing weight bearing patterns (14,17,18). Hence, differences will exist in pressure distribution and weight bearing through gait for two distinct individuals. Analysis of multiple factors in the human footprint should be expected to sample variations in bone geometry, muscle length, ligament tension and the gait cycle itself. Barefoot impressions represent a combination of multiple data inputs that may be used by experts to distinguish between them.

In our pilot study (5), a population of 960 Caucasian males was studied. Footprint impressions were collected from these volunteers using an inkless pad and specially-treated paper. These were analyzed to provide the probability distributions of the footprint measurements for the general population. A second group of 20 volunteers (11 male, 9 female) was measured nine times (three days over a one week period, three times each day) to provide a repetition sample. Using principal component analysis and standard deviation ratios of inter- and intra-population variances, the probability of a chance match was approximately  $10^{-8}$ . This repetition group of 20 individuals was too small a size to be an ideal random sample.

<sup>1</sup> Forensic Identification Research Services, Royal Canadian Mounted Police, Ottawa, Ontario K1A 0R2 Canada.

<sup>2</sup> Business Survey Methods Division, Statistics Canada, Ottawa, Ontario K1S 1A1 Canada.

<sup>3</sup> School of Mathematics and Statistics, Carleton University, Ottawa, Ontario K1S 5B6 Canada.

<sup>4</sup> The Orthopaedic Group, Clairton, PA 15025 and Robotics Institute, Dept. of Computer Science, Carnegie Mellon Univ., Pittsburgh, PA 15213.

\* This research was supported by grants from the Natural Sciences and Engineering Research Council (NSERC) of Canada and the Canadian Police Research Centre (CPRC).

Received 31 July 2004; and in revised form 23 April 2005; accepted 23 April 2005; published 3 Aug. 2005.



FIG. 1—A pair of barefoot impressions collected using the inkless pad.

Concerns about the repeated sample being a genuine cross-section of the population was one factor that led to this second, larger, corroboratory study. We study a larger population,  $n = 5755$ , in this follow-up work, with a repetition sample of size  $n = 134$ . We have improved our estimates of the statistical probability of a chance match. The general population was expanded by a factor of six, and now includes 47% females and 12% non-Caucasians.

### Materials and Methods

The same methodology for collection of the data was employed in this paper as in the pilot study (5). An inkless pad (Indicator) and chemically treated paper were used to collect the impressions. Each person was instructed to walk normally; to step onto the pad with one foot; then onto the floor with the other; then onto the paper with the next step of the inked foot; and to continue walking beyond the pad. This procedure happens to be compatible with the two-step gait initiation protocol (19). A practice run-through was performed to familiarize each subject with the process. The procedure was performed on a randomly chosen foot and then repeated with the other foot to produce a pair of impressions on each sheet of paper (See Fig. 1).

Each impression was scanned directly using an Epson GT10000 scanner set at 72 dpi. Each trace was then examined by a technician, under the supervision of the lead author, and any smudges found to be extraneous were removed using Corel PhotoPaint tools. On the rare occasions where the impression of the big toe was connected to the sole, a man-made channel was inserted by the technician using an erasure tool to give a straight one millimeter gap to ensure separation. A missing toe (e.g., if the smallest toe did not make contact with the paper) was represented by a single point, since the software rejected any scan that did not contain exactly 12 items: ten toes and two soles. Each trace was scanned and the data passed to an automated graphics program running under *AutoCAD* R13. The one significant change in the process employed in the current study

is that the tracing of the outline of each footprint impression was done by the *AutoCAD* software program, instead of being traced by a graphic artist.

The second phase in both studies was a repeated-measure experiment to determine the intra-personal (*individual*) variations of the footprint measurements. Three sets of barefoot impressions were collected daily from each subject over three days spread over a week. The magnitudes of the variations between different footprints of the same individual were contrasted to the interpersonal (*population*) variations.

### The Alignment of the Foot and the Measurements Taken

The convex hull of the trace of the sole of the foot was found and successive segments joining adjacent pairs of pixels of the convex hull found. The line segment of maximal length on each side of the sole was selected. The software excluded the middle third of every foot to correct for rare instances of exceptionally flat feet. This gave the correct answer for all normal feet and provided meaningful and consistent results for all footprints. This was also necessary for persons with high arches who produced faint or no markings in the middle third of their prints.

We found that there is always a significantly longest segment of the boundary of the cone on each side of the sole. These lines were projected to meet at the apex of an *enveloping cone* (5) below the heel. The *central axis* of the footprint outline was defined to be the bisector of the tangent cone, i.e., the line bisecting the cone angle through the apex. The outline of the footprint was rotated so that the central axis was aligned vertically with the apex at the bottom (See Fig. 2). The central axis has proven to be a stable concept, and other measurements depend on it.

The measurements were divided into five groups (See Fig. 3. See Table 1 for definitions):

- a) Foot measurements (lengths, widths, LD/RD, LE/RE, LF/RF, LG/RG, etc.)
- b) F-points (coordinates of points of metatarsal ridge, toes, etc.)
- c) L-points (widths of slices orthogonal to the axis, etc.)
- d) Areas (toes, soles.)
- e) Angles.

The foot measurements, in particular, were dependent on the alignment of the prints. The details of the measurements can be found in our earlier study (5), and are not repeated here. To summarize the process: the print was aligned so that the bisecting line was vertical, certain key points were determined (e.g., the center of the heel), and the magnitudes of these distances and angles were stored in a data base. The database was organized to include demographic data such as age, sex, race, height, and weight.

### The New Population and Repetition Samples

A major effort was made to collect a large set of prints from distinct individuals, both in Canada and the USA. Volunteer subjects were found in shopping centers, schools, training academies, etc. The goal was to build a large representative sample. Approximately half of the subjects were found in Ottawa, Ontario, Canada. However, since many measurements were done in public places, especially during the summer months, there were an unknown number of visitors from other locales. Almost 10% of the entire sample was collected in the USA. The remaining 40% was collected across Canada, with an emphasis on provinces other than Ontario.

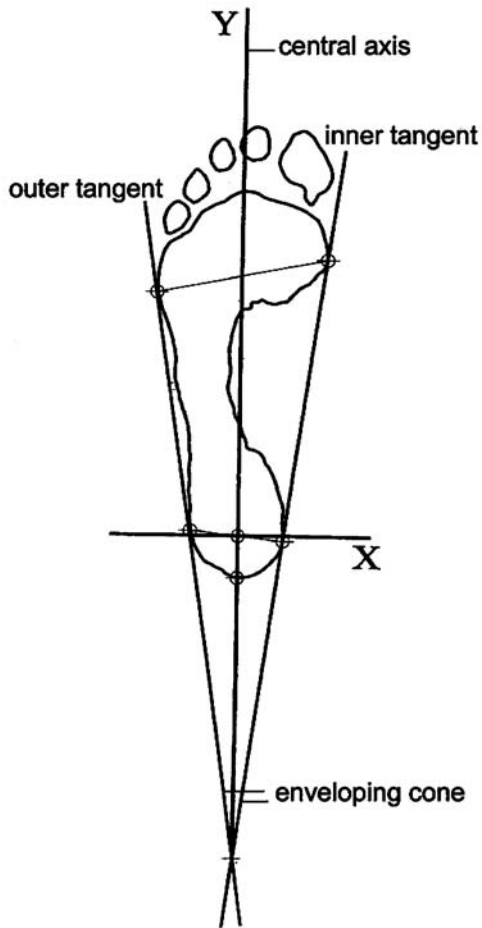


FIG. 2—Vertical alignment of the impression, illustrating the central axis.

The data were collected for each pair of feet. A dBASE<sup>©</sup> database was constructed to record and preserve the data for this and future studies. Table 2 contains three of many of the measurements gathered.

We received outstanding assistance from staff, faculty, and students at Carleton University during August 2003 in the collection of the repetition sample. The August timing of the data collection minimized the inconvenience to the subjects; many wore light summer footwear, and it was between the Summer and Fall terms. The repeated sample consists of 134 individuals: 60 males (49 Caucasian) and 74 females (64 Caucasian). The University population included persons from various ethnic origins. The gender-specific distributions of the sample show close resemblance to that of the large population samples, especially in terms of variability, which is the critical part of the study (see Tables 2, 3 and 4).

#### Statistical Analysis

This experiment has a full factorial design with two factors. First, 134 volunteer subjects provided three sets of barefoot impressions daily over a three-day period. The prints were scanned and the data passed to an automated graphics program running under AutoCAD R13. The prints were aligned vertically and measurements performed using exactly the same protocols as in our previous pilot study (5). In total,  $134 \times 3 \times 3$  sets of numerical measurements for each foot were taken in the repeated measurements phase. Approximately 200 measurements were gathered from every

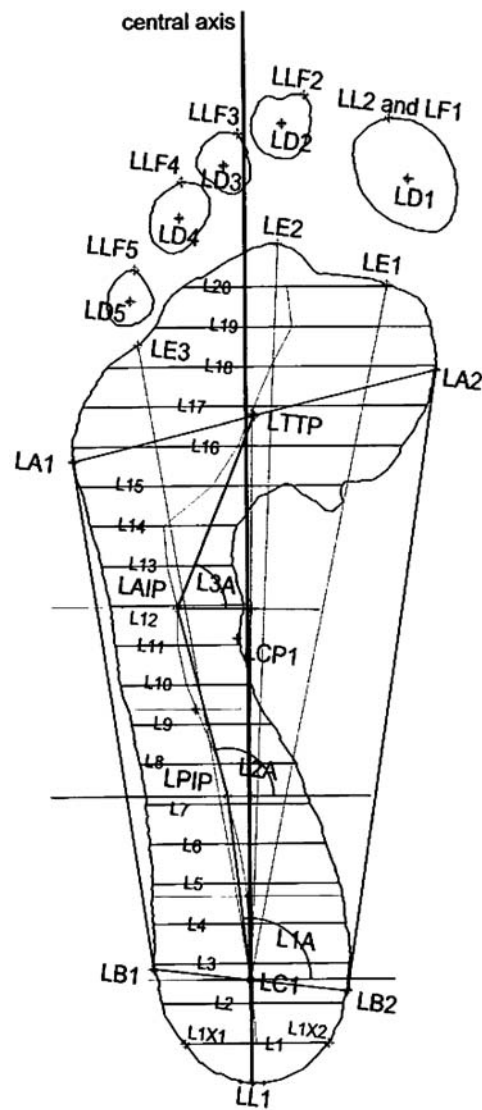


FIG. 3—Some of the many points and measurements taken on the barefoot impression.

barefoot impression; the database has 323 fields. Some measurements depend on the foot length (e.g., foot width at 1 cm. intervals), so some fields can be empty. About one-fifth of this data was utilized in the current study. The two statistical packages used for the analyses were SAS version 8 and SPSS 7.5.

These 134 repeated measure subjects were treated as a randomly selected sample of the general population. This permitted the analysis of the data as a two-way random-effects model (variance-components model) with two main variance components: the *personal effect* and the *daily effect*. For each quantitative measure, the variance of the personal effect,  $\sigma_p^2$ , represents the *inter-personal* (i.e., population) variance, and the variance of the daily effect,  $\sigma_d^2$ , is the *intra-personal* (i.e., individual) variance. These variance components are constructed from the three daily averages and are estimated from classic analysis of variance (ANOVA). The use of daily average for constructing variance components is analogous to the imprint worn inside a shoe, where one finds the pattern worn by the average of many impressions. The square roots of these two variance components give the corresponding standard deviations  $\sigma_p$  and  $\sigma_d$ . The standard deviations help determine the size of the

TABLE 1—Field names, their location, and definitions.

| LENGTH  | Length of Foot, from bottom of Heel to Tip of Toe giving the Greatest Measurement |
|---------|---|
| LBWIDTH | width of ball of foot   |
| LHWIDTH | width of heel   |
| LD1     | center of heel to center of 1st toe   |
| LD2     | center of heel to center of 2nd toe   |
| LD3     | center of heel to center of 3rd toe   |
| LD4     | center of heel to center of 4th toe   |
| LD5     | center of heel to center of 5th toe   |
| LE1     | bottom of heel to 1st metatarsal head   |
| LE2     | bottom of heel to 2nd metatarsal head   |
| LE3     | bottom of heel to 5th metatarsal head   |
| LF1     | bottom of heel to center of 1st toe   |
| LF2     | bottom of heel to center of 2nd toe   |
| LF3     | bottom of heel to center of 3rd toe   |
| LF4     | bottom of heel to center of 4th toe   |
| LF5     | bottom of heel to center of 5th toe   |
| LG1     | center of heel to 1st metatarsal head   |
| LG2     | center of heel to 2nd metatarsal head   |
| LG3     | center of heel to 5th metatarsal head   |
| LDBA1   | length of the outer side tangent to the foot                                      |
| LDBA2   | length of the inner side tangent to the foot                                      |
| L1A     | angle between horizontal and a line drawn between C1 and PIP                      |
| L2A     | angle between horizontal and a line drawn between AIP and PIP                     |
| L3A     | angle between horizontal and a line drawn between AIP and TTP                     |
| LABA1   | angle between horizontal and the outer tangent                                    |
| LABA2   | angle between horizontal and the inner tangent                                    |
| LD1A    | area of first toe   |
| LD1P    | perimeter of first toe  |
| LD2A    | area of second toe  |
| LD2P    | perimeter of second toe   |
| LD3A    | area of third toe   |
| LD3P    | perimeter of third toe  |
| LD4A    | area of fourth toe  |
| LD4P    | perimeter of fourth toe   |
| LD5A    | area of fifth toe   |
| LD5P    | perimeter of fifth toe  |
| LCENA   | area of foot minus the toes   |
| LCENP   | perimeter of foot minus the toes  |

TABLE 2—General population sample (n = 5755).\*

|                    | Male      | Caucasian | Age   | Height (cm) | Shoe Size |
|--------------------|-----------|-----------|-------|-------------|-----------|
| Mean or %          | 53.45     | 88.03     | 35.82 | 172.98      | 9.04      |
| Std. Dev.          |           |           | 11.93 | 10.41       | 1.78      |
| Males (n = 3076)   | Mean or % | 87.61     | 35.70 | 180.07      | 10.06     |
|                    | Std. dev. |           | 11.71 | 7.09        | 1.48      |
| Females (n = 2653) | Mean or % | 88.69     | 35.95 | 164.78      | 7.85      |
|                    | Std. dev. |           | 12.18 | 7.12        | 1.30      |

\* There was no sex recorded for a small number of subjects.

TABLE 3—Repetition population sample (n = 134).

|                  | Male      | Caucasian | Age   | Height (cm) | Shoe Size |
|------------------|-----------|-----------|-------|-------------|-----------|
| Mean or %        | 44.78     | 85.07     | 37.35 | 170.74      | 8.78      |
| Std. dev.        |           |           | 13.45 | 9.95        | 1.85      |
| Males (n = 60)   | Mean or % | 83.33     | 36.67 | 178.81      | 10.25     |
|                  | Std. dev. |           | 14.42 | 7.10        | 1.50      |
| Females (n = 74) | Mean or % | 86.49     | 37.88 | 164.21      | 7.71      |
|                  | Std. dev. |           | 12.73 | 6.54        | 1.26      |

TABLE 4—General population versus repetition sample: comparison of variances.

|         | Variable  | F statistic | p-value | Levene's Statistic | p-value |
|---------|-----------|-------------|---------|--------------------|---------|
| Males   | Height    | 0.0139      | 0.9061  | 0.0277             | 0.8677  |
|         | Shoe size | 1.1288      | 0.2881  | 1.1507             | 0.2835  |
| Females | Height    | 2.5681      | 0.1092  | 2.2941             | 0.1300  |
|         | Shoe size | 0.0134      | 0.9080  | 0.0115             | 0.9145  |

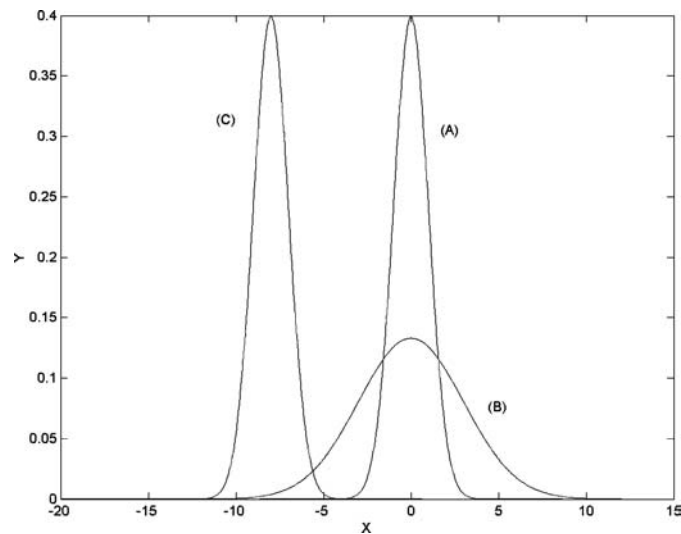


FIG. 4—Normal probability distribution curves. A) Normal distribution for a measurement for the entire population. B) Normal distribution taken from repeated measurements for an individual whose measure coincides with the average value for the population. The result of the intersection of curves A) and B) is relatively large, and represents the worst-case value given by equation (III). C) Normal distribution taken from repeated measurements for an individual whose measure coincides with a value in the wing of the normal distribution curve for the population. The chance match probability is represented by the intersection of curves A) and C), which is relatively small. The average case chance match probability is given by equation (IV). It is found by averaging this intersection across the entire population range.

tolerance window, or bin, and this size determines the chance match probability. The ratio of the two standard deviations,

$$\rho = \sigma_I / \sigma_P,$$

is called the *standard deviation ratio*. The standard deviation ratio gives the fraction of the total population distribution of a single measure, such as foot length, that is covered by the individual variation. The smaller the standard deviation ratio is, the more useful the measure and the more accurate a *match* of impressions that is made using this measure. A small standard deviation ratio  $\rho$  indicates a narrow overlap of the individual variation with that of the population (See Fig. 4).

### Results

Our major results include contributions to both the theoretical framework and the analysis of new experimental data. We consider the sensitive problem of statistical independence of data found in barefoot impressions for both single-dimensional and multiple-dimensional identification. Our aim is to provide a rigorous scientific foundation for the calculation of chance match probabilities with a mathematical guarantee of orthogonality of the variables and statistical independence.

### Chance Match Probability

One fundamental scientific challenge for modern forensic identification is the handling of chance match probability. For instance: how likely is it that a second unrelated person can match the forensic evidence collected from a crime scene or victim? In order to achieve the infinitesimal chance match probabilities that have been reported, modern forensic procedures use multiple characteristics or features of the evidence for identifying matches. In DNA typing (20), for example, these are either various alleles at a single locus, or genotypes of multiple loci. The statistical independence between alleles at a locus is sometimes addressed using Hardy-Weinberg Equilibrium, the condition in which allele frequencies are constant and inherited independently of one another. The statistical independence of genotypes between loci is called Linkage Equilibrium. To date, there is no empirical evidence that suggests either of these assumptions are violated in human populations. The slight deviations observed from Hardy Weinberg Equilibrium are adjusted with a  $\theta$  correction (21).

Let us first concentrate on matches for a single quantitative measure  $x$ . Once this measure is obtained from the individual subject, say  $x_s$ , a tolerance interval or bin,  $[x_s - \alpha\sigma_I, x_s + \alpha\sigma_I]$ , is created where  $\sigma_I$  is the individual standard deviation and  $\alpha$  is the bin size. We define  $\omega = \alpha\sigma_I$  and call  $2\omega$  the *bin width*. Note that  $\alpha\rho\sigma_P = \omega$ . A measurement  $x$  is deemed to be a *match* to the subject  $x_s$  in question if it falls inside the interval  $[x_s - \omega, x_s + \omega]$ .

Assume that in the general population the quantitative measure  $x$  has a probability distribution with a density function  $f(x)$ , then the chance match probability is given by

$$p(x_s, \omega) = \int_{x_s - \omega}^{x_s + \omega} f(w) dw. \quad (\text{I})$$

The Gaussian normal distribution function  $f(w) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(w-\mu)^2}{2\sigma^2}}$  is a major type of probability density distribution where  $\mu$  is the given mean, and  $\sigma$  the standard deviation. Its cumulative distribution function is closely related to the well-known *error function*,  $\text{erf}(x)$ , defined by

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt.$$

$$\text{Note that } \text{erf}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{2}x}^{\sqrt{2}x} e^{-t^2/2} dt.$$

The Gaussian normal distribution gives rise to the chance match probability function

$$P(x_s, \omega) = \frac{1}{\sqrt{2\pi}\sigma} \int_{x_s - \omega}^{x_s + \omega} e^{-\frac{(w-\mu)^2}{2\sigma^2}} dw \quad (\text{II})$$

that depends on both the individual measure and the bin width.

Theorem 1. Let the quantitative measure  $x_s$  in the general population follow a normal probability distribution with given (fixed) mean  $\mu_P$  and standard deviation  $\sigma_P$ .

- a) Fix the bin half-width at  $\omega$  and fix the individual variance  $\sigma_I$ . The maximum or worst-case chance match probability is achieved when the subject measure  $x_s$  coincides with the population mean,  $x_s = \mu_P$ . The maximum (or worst-case) probability is then given by

$$P_{\max}(\omega) = \frac{1}{\sqrt{2\pi}\sigma_P} \int_{\mu_P - \omega}^{\mu_P + \omega} e^{-\frac{(w-\mu_P)^2}{2\sigma_P^2}} dw = \text{erf}(\sqrt{2}\alpha\rho/2) \quad (\text{III})$$

- b) The average, or expected, chance match probability for a subject randomly chosen from the general population is given

TABLE 5—Worst-case and average chance match probabilities using a single measure.\*

| Standard Deviation Ratio | Worst-case Chance Match Probability | Average Chance Match Probability | Ratio: Avg Case $\div$ Worst Case |
|--------------------------|-------------------------------------|----------------------------------|-----------------------------------|
| 1:2                      | 0.3829                              | 0.2763                           | 0.7216202                         |
| 1:4                      | 0.1974                              | 0.1403                           | 0.7107761                         |
| 1:10                     | 0.0797                              | 0.0564                           | 0.7076956                         |
| 1:20                     | 0.0399                              | 0.0282                           | 0.7072540                         |
| 1:40                     | 0.0199                              | 0.0141                           | 0.7071436                         |
| 1:60                     | 0.0133                              | 0.0094                           | 0.7071231                         |
| 1:80                     | 0.0100                              | 0.0071                           | 0.7071159                         |
| 1:100                    | 0.0080                              | 0.0056                           | 0.7071126                         |

\* In this table we assume that  $a = 1$ . Hence the worst case chance probability is  $\text{erf}(\rho/\sqrt{2})$  and the average chance match probability is  $\text{erf}(\rho/2)$ .

by

$$\begin{aligned} \bar{P}(\omega) &= E(P(x, \omega)) = \frac{1}{2\pi\sigma_P^2} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu_P)^2}{2\sigma_P^2}} \int_{x-\omega}^{x+\omega} e^{-\frac{(y-\mu_P)^2}{2\sigma_P^2}} dy dx \\ &= \text{erf}(\alpha\rho/2) \end{aligned} \quad (\text{IV})$$

$$\text{c) } \lim_{\omega \rightarrow 0} \frac{\bar{P}(\omega)}{P_{\max}(\omega)} = \frac{\sqrt{2}}{2} = 0.70710678.$$

The proof of Theorem 1 is found in Appendix 1. The ratio of  $P_{\max}$  to  $\bar{P}$  stabilizes quickly; e.g., when the half-width  $\omega = \sigma_I$ , the ratio is 0.70725 for  $\rho = 1/20$  and 0.70714 for  $\rho = 1/40$  (See Table 5).

Equation (III) is intuitively clear: when the subject's measure coincides with the population mean, a large percentage of the population may have a measure close enough to that of the individual to fall into the tolerance window. This represents a *worst-case scenario*. When  $x_s$ , the subject's measure, moves away from the population mean  $\mu_P$ , the proportion of the population members who have similar measures decreases, as can be seen in Figure 3, and as given in Equation (IV).

It is extremely unlikely to have a subject all of whose individual measures coincide with the corresponding population mean in all dimensions. When multiple measures are used simultaneously for a footprint comparison, especially if these measures are statistically independent of each other, we cannot expect *on average* to be in the situation of Equation (III). Therefore, for comparisons using multiple measures, it is more appropriate to use the average chance match probability given by Equation (IV). Both the worst-case and average chance match probabilities depend only on the tolerance window size  $2\omega$  and the standard deviation ratio  $\rho$ .

Furthermore, it is imperative to note that the chance match probabilities for statistically independent dimensions can be multiplied to obtain the overall chance match probability. In our method, the multiple measures we use for comparison are eigenvectors constructed via a principal component analysis. These always are statistically independent. As a result, our probability estimates are based on mathematical theorems, and not on hypotheses that are extremely difficult to verify or technical assumptions that might conflict with observations.

### Normality of the Data

It has been widely observed that a number of quantifiable human characteristics, e.g., weight and height, follow a normal probability distribution in a general population. In the pilot study (5), 38 quantitative measurements derived from barefoot impressions

TABLE 6—Normality of principal components.\*

|    | Females                      | $n = 2653$ | Males                        | $n = 3076$ |
|----|------------------------------|------------|------------------------------|------------|
| PC | Kolmogorov-Smirnov Statistic | $p$ -value | Kolmogorov-Smirnov Statistic | $p$ -value |
| 1  | 0.012434                     | >0.1500    | 0.013304                     | >0.1500    |
| 2  | 0.010478                     | >0.1500    | 0.011220                     | >0.1500    |
| 3  | 0.028236                     | 0.0100     | 0.016690                     | 0.1053     |
| 4  | 0.011587                     | >0.1500    | 0.014196                     | >0.1500    |
| 5  | 0.015596                     | >0.1500    | 0.015651                     | 0.1166     |
| 6  | 0.014936                     | >0.1500    | 0.012477                     | >0.1500    |

\* A  $p$ -value greater than 0.05 for the Kolmogorov-Smirnov statistic means the distribution of the variable is not significantly different from a normal distribution. We show the first 6 components but only use the first 5.

were selected. The normality of most of this data from each foot was verified in a large database consisting of 960 Caucasian males. Many other variables and control measurements compiled in the current phase are not used here because we found that these 38 alone were sufficient to achieve our primary goal. In our much expanded new databases, this normality is confirmed again as most of the collected quantitative measurements closely follow the normal distribution. The normal probability plots of variables LLENGTH and LBWIDTH in Fig. 5 demonstrate this close fit.

There are two additional arguments that support the use of normal distributions for calculating the chance match probabilities. First, as has already been noted, we use principal components for our final probability calculation. These principal components are weighted sums of a large number (38 in most cases) of measures many of which themselves closely follow a normal distribution. It is well known that the Central Limit Theorem often holds for these sums of non-independent random variables when certain regularity conditions are met (22). Second, as in many actual applications in other fields, violations of normality occasionally arise when extreme values or *outliers* occur. However, this is not likely to cause problems in either theory or practice.

More specifically, extreme values would in theory lead to very small chance match probabilities by the first part of Theorem 1. As the attached Fig. 4 illustrates, individual probability distributions centered about these extreme personal values would have minimal overlap with the general population probability distribution, leading to very small chance-match probabilities. In practice, an impression with extreme values far off from the population mean is observably distinct from the common population and hard to mismatch in practice. In addition, extreme values can often be data errors in the databases that are purged during data cleaning.

As Table 6 and sample graphs (See Fig. 5) demonstrate, normality holds for most principal components constructed from the original measurements. The only strong exception is the third principal component in the female data. Even in this case, the normal probability shows a fairly good fit to the normal distribution. In the normal quantile plot, one sees that the lack of fit occurs only in the extreme tail ranges.

#### The Repeated-Measure Experiment

The repeated measure experiment is a central part of our study. Table 7 gives a summary of the estimated individual and population variances, and their ratios, for 38 selected measurements. The ratio of the estimated individual and population standard deviations are tabulated for convenience. Our population variance estimates are

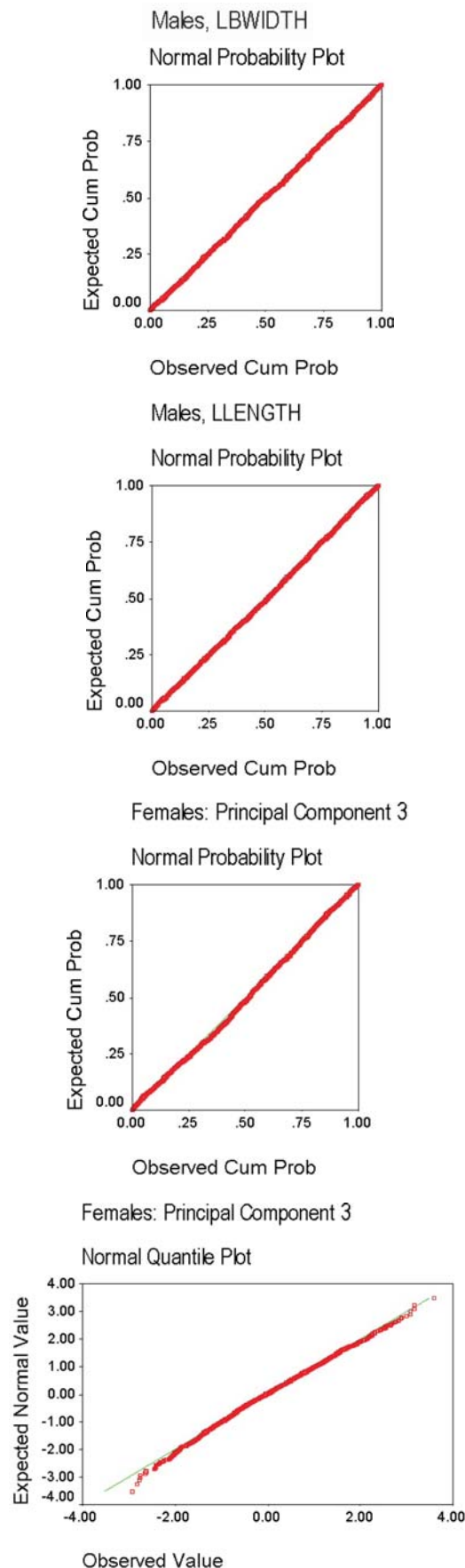


FIG. 5—Examples of normal probability plots for two of the original measurements, and for two examples of the principal components.

TABLE 7—Individual and population variances.

| Variable Name | Estimated Population Mean | Estimated Population Variance $\sigma_p^2$ | Estimated Individual Variance $\sigma_I^2$ | Estimated Std. Dev. Ratio $\sigma_I/\sigma_p$ |
|---------------|---------------------------|--|--|---|
| LLENGTH       | 246.01                    | 365.74                                     | 0.0699                                     | 0.013827                                      |
| LBWIDTH       | 89.23                     | 54.10                                      | 0.0137                                     | 0.015900                                      |
| LHWIDTH       | 48.91                     | 28.79                                      | 0.0762                                     | 0.051440                                      |
| LD1           | 199.93                    | 270.77                                     | 0.0570                                     | 0.014505                                      |
| LD2           | 206.96                    | 307.97                                     | 0.0377                                     | 0.011065                                      |
| LD3           | 197.98                    | 282.02                                     | 0.0845                                     | 0.017310                                      |
| LD4           | 185.54                    | 238.22                                     | 0.0540                                     | 0.015055                                      |
| LD5           | 169.33                    | 191.33                                     | 0.0000                                     | 0.010001*                                     |
| LE1           | 202.38                    | 245.33                                     | 0.0675                                     | 0.016580                                      |
| LE2           | 204.00                    | 231.85                                     | 0.0375                                     | 0.012724                                      |
| LE3           | 182.54                    | 205.24                                     | 0.0000                                     | 0.010001                                      |
| LF1           | 245.10                    | 366.52                                     | 0.0636                                     | 0.013171                                      |
| LF2           | 242.24                    | 372.53                                     | 0.0241                                     | 0.008036                                      |
| LF3           | 232.77                    | 341.71                                     | 0.0798                                     | 0.015281                                      |
| LF4           | 219.68                    | 289.59                                     | 0.0820                                     | 0.016826                                      |
| LF5           | 202.07                    | 245.28                                     | 0.0000                                     | 0.010001                                      |
| LG1           | 176.52                    | 207.96                                     | 0.0735                                     | 0.018800                                      |
| LG2           | 177.69                    | 194.54                                     | 0.0402                                     | 0.014374                                      |
| LG3           | 156.63                    | 172.55                                     | 0.0000                                     | 0.010001                                      |
| RLENGTH       | 246.10                    | 376.00                                     | 0.1392                                     | 0.019238                                      |
| RBWIDTH       | 89.32                     | 55.49                                      | 0.0080                                     | 0.012017                                      |
| RHWIDTH       | 49.17                     | 28.69                                      | 0.0457                                     | 0.039936                                      |
| RD1           | 199.67                    | 277.81                                     | 0.1114                                     | 0.020024                                      |
| RD2           | 206.72                    | 313.93                                     | 0.1357                                     | 0.020794                                      |
| RD3           | 197.48                    | 287.34                                     | 0.0702                                     | 0.015634                                      |
| RD4           | 185.22                    | 241.06                                     | 0.0471                                     | 0.013972                                      |
| RD5           | 168.90                    | 196.44                                     | 0.0249                                     | 0.011267                                      |
| RE1           | 202.59                    | 248.52                                     | 0.3994                                     | 0.040080                                      |
| RE2           | 204.25                    | 237.13                                     | 0.2321                                     | 0.031289                                      |
| RE3           | 182.78                    | 205.72                                     | 0.0981                                     | 0.021838                                      |
| RF1           | 245.13                    | 376.99                                     | 0.1692                                     | 0.021181                                      |
| RF2           | 242.45                    | 378.33                                     | 0.1422                                     | 0.019391                                      |
| RF3           | 232.75                    | 343.96                                     | 0.1711                                     | 0.022301                                      |
| RF4           | 219.66                    | 290.18                                     | 0.1473                                     | 0.022532                                      |
| RF5           | 202.37                    | 245.62                                     | 0.0792                                     | 0.017956                                      |
| RG1           | 176.25                    | 212.75                                     | 0.2460                                     | 0.034002                                      |
| RG2           | 177.46                    | 201.10                                     | 0.0790                                     | 0.019825                                      |
| RG3           | 156.41                    | 176.53                                     | 0.0000                                     | 0.010001                                      |

\* The zero value here and below would normally give an estimated standard deviation ratio of zero. This is replaced here and below with the conservative value  $100/9999 = .0100010001\dots$

taken from the large population sample, not the relatively small repetition sample. This makes the estimated standard deviation ratio a great deal more representative and stable.

#### Application of the Standard Deviation Ratio

Chance matches of measurements of a pair of barefoot impressions are dependent on the number of variables used and the standard deviation ratio of each variable. For instance, from Table 5, measurements with an estimated standard deviation ratio of 60 each provide a worstcase chance match probability of 1.33%. These variables may be statistically correlated with each other, so we do not multiply the individual chance match probabilities when multiple measurements are used for identification. However, principal component analysis provides statistically independent measurements corresponding to new, rotated, statistically independent variables derived from our set of 38 variables. Probabilities corresponding to independent variables can be multiplied.

A principal component analysis was performed on 38 measurements on the barefoot impressions. We considered four populations: combined; male; female; and Caucasian men. We decided to

TABLE 8—Principal component analysis.

| Component Number                                      | Sum of Squared Loadings | % of Variance Explained | Cumulative % of Variance | Estimated Standard Deviation Ratio $\rho$ |
|---|-------------------------|-------------------------|--------------------------|---|
| a. Principal component analysis (Combined population) |                         |                         |                          |   |
| 1   | 12.61                   | 33.17                   | 33.17                    | 0.010001                                  |
| 2   | 12.05                   | 31.70                   | 64.87                    | 0.016458                                  |
| 3   | 5.22                    | 13.73                   | 78.61                    | 0.061766                                  |
| 4   | 2.88                    | 7.57                    | 86.17                    | 0.063171                                  |
| 5   | 2.78                    | 7.31                    | 93.49                    | 0.021491                                  |
| b. Principal component analysis (females only)        |                         |                         |                          |   |
| 1   | 9.50                    | 24.99                   | 24.99                    | 0.010001                                  |
| 2   | 9.45                    | 24.87                   | 49.86                    | 0.010001                                  |
| 3   | 8.67                    | 22.82                   | 72.68                    | 0.060864                                  |
| 4   | 3.06                    | 8.06                    | 80.74                    | 0.082169                                  |
| 5   | 2.31                    | 6.07                    | 86.81                    | 0.010612                                  |
| c. Principal component analysis (males only)          |                         |                         |                          |   |
| 1   | 10.12                   | 26.62                   | 26.62                    | 0.010001                                  |
| 2   | 9.21                    | 24.23                   | 50.85                    | 0.011960                                  |
| 3   | 9.13                    | 24.02                   | 74.87                    | 0.021925                                  |
| 4   | 2.82                    | 7.41                    | 82.28                    | 0.044762                                  |
| 5   | 2.43                    | 6.41                    | 88.69                    | 0.021949                                  |
| d. Principal component analysis (Caucasian males)     |                         |                         |                          |   |
| 1   | 10.22                   | 26.90                   | 26.90                    | 0.010001                                  |
| 2   | 9.12                    | 24.01                   | 50.92                    | 0.039510                                  |
| 3   | 8.70                    | 22.89                   | 73.81                    | 0.019398                                  |
| 4   | 2.84                    | 7.47                    | 81.28                    | 0.047058                                  |
| 5   | 2.40                    | 6.31                    | 87.59                    | 0.010001                                  |

use only first five principal components corresponding to orthogonal eigenvectors, given that each component explains a substantive proportion of the variability in the data. They were extracted to give five new independent variables in each of the four populations.

In the remainder of this article we choose a window or bin width equal to two intra-personal standard deviations or  $2\sigma_I$ . This is equivalent to taking  $\alpha = 1$  in the notations leading to our Theorem 1. This choice is prompted by two practical aspects of barefoot impressions, namely an abundant, low-cost supply of comparison samples, and relatively large variability. In any event, our mathematical framework holds true irrespective of the bin size.

Table 8 illustrates that the first five principal components account for approximately 93.5% of the total variance of the 38 variables in the case of the combined populations. The fifth component accounts for at least 6% of the variance of the measurements of the footprint impressions in each of the four populations. Table 8 gives separate results for the four populations.

In the average case, we calculate the average based on normally distributed measurements overlapping anywhere on the curve for the entire population. For the combined population (See Table 8), the average probability of a chance match is  $7.88 \times 10^{-10}$ . The latter is equivalent to stating that the odds of a chance match are 1:1,268,000,000 or 1 in 1.27 billion. The odds change for the various populations. The cumulative worst-case and average chance match probabilities are calculated in Table 9 based on Theorem 1 and the statistical independence of the principal components. These probabilities are derived from the estimated standard deviation ratio given in the last column of Table 9, and these estimated average case probabilities of a chance match and the odds of chance match are found in Table 10.

The worst case deals with the eventuality where the subject has the mean values as his measurements in every dimension; hence the probability of a chance-match is higher. This gives a conservative estimate of the chance match probability (II) and offers a good

TABLE 9—Cumulative worst case and average probabilities of a match.\*

| Component Number                                      | Cumulative % of Variance | Estimated Standard Deviation Ratio $\rho$ | Cumulative Worst Case Probability of a Match | Cumulative Average Case Probability of a Match |
|---|--------------------------|---|--|--|
| a. Principal component analysis (Combined population) |                          |   |  |  |
| 1   | 33.17                    | 0.0100                                    | $7.9795 \times 10^{-3}$                      | $5.6424 \times 10^{-3}$                        |
| 2   | 64.87                    | 0.0165                                    | $1.0478 \times 10^{-4}$                      | $5.2392 \times 10^{-5}$                        |
| 3   | 78.61                    | 0.0618                                    | $5.1606 \times 10^{-6}$                      | $1.8252 \times 10^{-6}$                        |
| 4   | 86.17                    | 0.0632                                    | $2.5994 \times 10^{-7}$                      | $6.5028 \times 10^{-8}$                        |
| 5   | 93.49                    | 0.0215                                    | $4.4570 \times 10^{-9}$                      | $7.8846 \times 10^{-10}$                       |
| b. Principal component analysis (females only)        |                          |   |  |  |
| 1   | 24.99                    | 0.0100                                    | $7.9795 \times 10^{-3}$                      | $5.6424 \times 10^{-3}$                        |
| 2   | 49.86                    | 0.0100                                    | $6.3673 \times 10^{-5}$                      | $3.1837 \times 10^{-5}$                        |
| 3   | 72.68                    | 0.0609                                    | $3.0902 \times 10^{-6}$                      | $1.0929 \times 10^{-6}$                        |
| 4   | 80.74                    | 0.0822                                    | $2.0237 \times 10^{-7}$                      | $5.0638 \times 10^{-8}$                        |
| 5   | 86.81                    | 0.0106                                    | $1.7135 \times 10^{-9}$                      | $3.0318 \times 10^{-10}$                       |
| c. Principal component analysis (males only)          |                          |   |  |  |
| 1   | 26.62                    | 0.0100                                    | $7.9795 \times 10^{-3}$                      | $5.6424 \times 10^{-3}$                        |
| 2   | 50.85                    | 0.0120                                    | $7.6146 \times 10^{-5}$                      | $3.8074 \times 10^{-5}$                        |
| 3   | 74.87                    | 0.0219                                    | $1.3320 \times 10^{-6}$                      | $4.7095 \times 10^{-7}$                        |
| 4   | 82.28                    | 0.0448                                    | $4.7556 \times 10^{-8}$                      | $1.1892 \times 10^{-8}$                        |
| 5   | 88.69                    | 0.0219                                    | $8.3277 \times 10^{-10}$                     | $1.4725 \times 10^{-10}$                       |
| d. Principal component analysis (Caucasian males)     |                          |   |  |  |
| 1   | 26.90                    | 0.0100                                    | $7.9795 \times 10^{-3}$                      | $5.6424 \times 10^{-3}$                        |
| 2   | 50.92                    | 0.0395                                    | $2.5148 \times 10^{-4}$                      | $1.2576 \times 10^{-4}$                        |
| 3   | 73.81                    | 0.0194                                    | $3.8922 \times 10^{-6}$                      | $1.3763 \times 10^{-6}$                        |
| 4   | 81.28                    | 0.0471                                    | $1.4609 \times 10^{-7}$                      | $3.6535 \times 10^{-8}$                        |
| 5   | 87.59                    | 0.0100                                    | $1.1657 \times 10^{-9}$                      | $2.0615 \times 10^{-10}$                       |

\* The  $k$ -th row of the last two columns are the products  $\prod_{i=1}^k \text{erf}(\frac{\rho_i}{\sqrt{2}})$  and  $\prod_{i=1}^k \text{erf}(\frac{\rho_i}{2})$  respectively.

TABLE 10—Odds of a chance match for four populations.

| Population Group    | Probability            | Odds of a Chance Match |
|---------------------|------------------------|------------------------|
| Combined Population | $7.88 \times 10^{-10}$ | $1:1.27 \times 10^9$   |
| Females             | $3.03 \times 10^{-10}$ | $1:3.30 \times 10^9$   |
| Males               | $1.47 \times 10^{-10}$ | $1:6.79 \times 10^9$   |
| Caucasian Males     | $2.06 \times 10^{-10}$ | $1:4.85 \times 10^9$   |

initial indication of the uniqueness of barefoot impressions. In the pilot study, we only provided the worst case results. Theorem 1 finds the average chance match probability taken over the entire range of the population (III). This gives a more precise and representative result compared to the worst case analysis.<sup>5</sup>

## Discussion

This study is an enhancement of the earlier pilot study. We repeated the basic experiment but increased the sample size from 960 to 5755 and automated a crucial part of the data analysis. We improved the heterogeneity of the population sample and added subjects of mixed racial and national background. The repeated sample proved to be an excellent sample of our total population. A more rigorous and general mathematical framework for calculating chance match probabilities was developed, whose application is not limited to barefoot impressions.

<sup>5</sup> We also performed a Likelihood Ratio approach to studying the barefoot impression data. We used the multivariate normal (MVN) random-effect model introduced by Aitken and Lucy (23). We found that the likelihood ratio approach led to similar observations about uniqueness as expressed above and provided no further advantage.

Our mathematical contributions include Theorem 1, where we determine the average chance match probabilities and find their ultimate relationship to the worst-case probabilities. We show that population means do not affect chance match probabilities, whereas population variances have a direct impact. It is noteworthy that in terms of variance, our repetition sample and general population sample are fairly close. Based on both Levene's test and the F test for comparing the variance of the two samples, all differences are statistically insignificant (See Table 4). On the other hand, the differences in mean height are due to a relatively larger female and Asian component in the repetition sample than in the population sample. These means, however, have no bearing to either variance or eigenvectors critical to our study.

Another mathematical contribution of this paper is the application of principal components as the appropriate research methodology here. This guarantees statistical independence of the components and trumps any objection to the use of the product rule. This gives a practical option to avoid the predicament that other statistical approaches have encountered. We use only the first five components here to avoid the danger of using spurious eigenvectors<sup>6</sup>, and in concordance with our moderate approach to the calculations.

Previously, the outlines of the footprint impressions were manually traced by a graphic artist onto celluloid sheets, and these traces were then scanned. This process was time-consuming and may have introduced an additional source of variation. This was replaced by a software program to determine the outline directly from a scan of the actual impression, with very little human intervention.

<sup>6</sup> Based on the rotated solutions, for all four populations under study, we had five or more eigenvalues greater than one (the Kaiser criterion). In addition, each of these eigenvectors explained more than 5% of the total variance in the measurement data.



From that point onward, all other methods of data collection were identical to the pilot study.

Our approach was based on the comparison of personal to the population variation, so accurate estimation of these variance components was critical. The population variation estimates used in the pilot study were based only on the small repeated sample, whose representation of the general population could not be directly verified due to data compatibility problems. This was resolved here. In this study, the repeated sample had inter-personal variations very close to those of the large population database in critical demographic variables (height and shoe size) and most foot measures. All population variance components used in the final comparison came from the large population database, so the new standard deviation ratio estimates were highly reliable. Further, all eigenvalues and eigenvectors were constructed from the combined large database, making the constructions more stable and robust.

In further studies of the shape of footprints, there is a sizeable amount of additional data that can be generated from the inked impressions collected in our database. It is possible to further analyze these barefoot impressions by introducing more variables, especially non-linear ones, to produce tighter bounds on the chance match probability. The use of approximations of the shape and landmarks (24) provides a rich set of opportunities. We may be able to generalize our methodology to other human physical characteristics that can be similarly quantified.

#### Acknowledgments

We wish to thank Peter H. Petersen for his many helpful observations and hard work in the pilot study and early data entry. We want to express our appreciation to our research librarian, Nerine Waldron, for her outstanding assistance. J.D. Schneider did much of the database computer programming. We are grateful to Professor George Carmody of Carleton University for sharing his deep knowledge of DNA with us. Rick Arnold and Mariana Lafrance, who collected the key Carleton footprint data with enthusiasm and care, merit special note. Valerie Hawkins and Adam Miller provided assistance with data entry. We also want to express our sincere gratitude to the thousands of volunteers who donated their barefoot impressions, and all the research assistants who helped to collect them.

#### APPENDIX 1—Proof of Theorem 1

**Proof:** a) It is easily confirmed that setting the partial derivative of the chance match probability (2) with respect to  $x_s$  equal to 0

$$\frac{\partial P}{\partial x_s} = \frac{1}{\sqrt{2\pi}\sigma_p} \left( e^{-\frac{(x_s + \omega - \mu_p)^2}{2\sigma_p^2}} - e^{-\frac{(x_s - \omega - \mu_p)^2}{2\sigma_p^2}} \right) = 0$$

gives  $x_s = \mu_p$ . Moreover, the 2nd derivative  $\frac{\partial^2 P(x_s, \omega)}{\partial x_s^2} < 0$  at  $x_s = \mu_p$ . Therefore  $P(x_s, \omega)$ , viewed as a function of a single variable  $x_s$ , is maximized on each  $\omega$  cross-section at  $x_s = \mu_p$ . In other words, the maximum probability (or worst case) corresponds to the situation when the subject mean is equal to the population mean (see Fig. 5). Substitute  $y = (w - \mu_p)/\sigma_p$ , to obtain

$$P_{\max}(\omega) = \frac{1}{\sqrt{2\pi}\sigma_p} \int_{\mu_p - \omega}^{\mu_p + \omega} e^{-\frac{(w - \mu_p)^2}{2\sigma_p^2}} dw$$

$$d\omega = \frac{1}{\sqrt{2\pi}} \int_{-\alpha\rho}^{\alpha\rho} e^{-\frac{y^2}{2}} dy = \text{erf}(\sqrt{2}\alpha\rho/2).$$

b) We assume that the subject's measure follows the general population distribution. Then, the chance match probability that another individual *unrelated* to the subject is a match to the first is expressed as a double integral:

$$\bar{P}(v) = E(P(x, v)) = \frac{1}{2\pi\sigma_p^2} \int_{-\infty}^{\infty} e^{-\frac{(x - \mu_p)^2}{2\sigma_p^2}} \int_{x-v}^{x+v} e^{-\frac{(y - \mu_p)^2}{2\sigma_p^2}} dy dx.$$

Note that  $\bar{P}(0) = 0$  and

$$\begin{aligned} \frac{d\bar{P}(v)}{dv} &= \frac{1}{2\pi\sigma_p^2} \int_{-\infty}^{\infty} e^{-\frac{(x - \mu_p)^2}{2\sigma_p^2}} \frac{d}{dv} \int_{x-v}^{x+v} e^{-\frac{(y - \mu_p)^2}{2\sigma_p^2}} dy dx \\ &= \frac{1}{2\pi\sigma_p^2} \int_{-\infty}^{\infty} e^{-\frac{(x - \mu_p)^2}{2\sigma_p^2}} \left( e^{-\frac{(x - \mu_p + v)^2}{2\sigma_p^2}} + e^{-\frac{(x - \mu_p - v)^2}{2\sigma_p^2}} \right) dx \\ &= \frac{1}{2\pi\sigma_p^2} e^{-\frac{v^2}{4\sigma_p^2}} \int_{-\infty}^{\infty} \left( e^{-(x - \mu_p + \frac{v}{2})^2/\sigma_p^2} + e^{-(x - \mu_p - \frac{v}{2})^2/\sigma_p^2} \right) dx. \end{aligned}$$

Substitute twice:  $u = (x - \mu_p + \frac{v}{2})/\sigma_p$ ,  $u = (x - \mu_p - \frac{v}{2})/\sigma_p$ , to find

$$\frac{d\bar{P}(v)}{dv} = \frac{1}{\pi\sigma_p} e^{-\frac{v^2}{4\sigma_p^2}} \int_{-\infty}^{\infty} e^{-u^2} du = \frac{1}{\sqrt{\pi}\sigma_p} e^{-\frac{v^2}{4\sigma_p^2}}.$$

Therefore,

$$\bar{P}(\omega) = \bar{P}(0) + \int_0^{\omega} \frac{d\bar{P}}{dv} dv = \frac{1}{\sqrt{\pi}\sigma_p} \int_0^{\omega} e^{-\frac{v^2}{4\sigma_p^2}} dv.$$

Substitute  $t = v/(\sqrt{2}\sigma_p)$  and use the symmetry of the integrand, to obtain the final formula:

$$\bar{P}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\alpha\rho/\sqrt{2}}^{\alpha\rho/\sqrt{2}} e^{-\frac{t^2}{2}} dt = \text{erf}(\alpha\rho/2).$$

c) The proof of this part follows from a single application of L'Hôpital's rule

$$\begin{aligned} \lim_{\omega \rightarrow 0} \frac{\text{erf}(\alpha\rho/2)}{\text{erf}(\alpha\rho/2)} &= \frac{\lim_{\omega \rightarrow 0} \frac{2}{\sqrt{\pi}} \int_0^{\omega/2\sigma_p} e^{-t^2} dt}{\lim_{\omega \rightarrow 0} \frac{2}{\sqrt{\pi}} \int_0^{\omega/\sqrt{2}\sigma_p} e^{-t^2} dt} \\ &= \frac{\lim_{\omega \rightarrow 0} \frac{2}{2\sigma_p\sqrt{\pi}} e^{-\omega^2/4\sigma_p^2}}{\lim_{\omega \rightarrow 0} \frac{2}{\sqrt{2}\sigma_p\sqrt{\pi}} e^{-\omega^2/2\sigma_p^2}} = \frac{\sqrt{2}}{2}. \end{aligned}$$

#### References

- Kennedy RB. Uniqueness of bare feet and its use as a possible means of identification. *Forensic Sci Int* 1996;82:81–7. [PubMed]
- Kennedy RB. Preliminary study on the uniqueness of barefoot impressions. *Can Soc Forensic Sci J* 1996;29(4):233–8.
- Bodziak WJ. Footwear impression evidence. Boca Raton: CRC Press, 2000;381–411.
- Kennedy RB. Bare footprint marks. In: Siegel JA, Saukko PJ, Knupfer GC, editors. *Encyclopedia of Forensic Sciences*, Volume 3. London: Academic Press, 2000; 1189–95.
- Kennedy RB, Pressman IS, Chen S, Pressman AE, Petersen PH. Statistical analysis of barefoot impressions. *J Forensic Sci* 2003;48(1):55–63. [PubMed]
- Kennedy RB, Gehl S, Massey S, Saunders G. Use of barefoot morphology in criminal investigations. *Identification Canada* 2002;25(4):4–6.
- Robbins LM. The individuality of human footprints. *J Forensic Sci* 1978;23:778–85. [PubMed]
- Qamra SR, Sharma BR, Kaila P. Naked foot marks – A preliminary study of identification factors. *Forensic Sci Int* 1980;16:145–52. [PubMed]

9. Laskowski GE, Kyle VL. Barefoot impressions – A preliminary study of identification characteristics and population frequency and their morphological features. *J Forensic Sci* 1988;33(2):378–88. [\[PubMed\]](#)
10. Barker SL, Scheuer JL. Predictive value of human footprints in a forensic context. *Med Sci Law* 1998;38:341–6. [\[PubMed\]](#)
11. Mann RA, editor. *DuVries surgery of the foot*. 4th ed. St. Louis: CV Mosby Co., 1978.
12. Taylor AJ, Menz HB, Keenan A-M. [The influence of walking speed on plantar pressure measurement using the two-step gait initiation protocol](#). *The foot* 2004;14:49–55.
13. Elftman HO. [A cinematic study of the distribution of pressure in the human foot](#). *Anat Record* 1934;59:481–7.
14. Grundy M, Blackburn G, Tosh PA, McLeish RD, Smidt L. An investigation of the centers of pressure under the foot while walking. *J Bone Surg Br* 1975;57:98–103.
15. Hennig EM, Rosenbaum D. Pressure distribution patterns under the feet of children in comparison with adults. *Foot & Ankle* 1991;11(5):306–11.
16. Hutton WC, Stott JRR, Stokes IAF. Methods of studying the mechanics of the foot. In: Klenerman LK, editor. *The foot and its disorders*. London: Blackwell, 1976.
17. Beverly MC, Horan FT, Hutton WC. [Load cell analysis following silastic arthroplasty of the hallux](#). *International Orthopaedics* 1985;9:101–4. [\[PubMed\]](#)
18. Pertunnen J. Foot loading in normal and pathological walking. Studies in sport, physical education and health [dissertation]. Finland: University of Jyväskylä, 2002.
19. Meyers-Rice B, Sugars L, McPoil T, Cornwall MW. Comparison of three methods for obtaining plantar pressures in nonpathologic subjects. *J Am Podiatr Med Assoc* 1994;84:499–504. [\[PubMed\]](#)
20. Evett I, Weir B. *Interpreting DNA evidence: Statistical genetics for forensic scientists*. Sunderland, Mass: Sinauer Associates, 1998.
21. U.S. National Research Council, Committee on DNA Forensic Science. *Evaluation of forensic DNA evidence*. Washington, DC: National Academy Press, 1996.
22. Dvoretzky A. Asymptotic normality for sums of dependent random variables. *Proceedings Sixth Berkeley Symposium on Statistics and Probability*. Berkeley, CA: University of California Press, 1972;513–35.
23. Aitken CGG, Lucy, D. [Evaluation of trace evidence in the form of multivariate data](#). *J Royal Stat Soc Series C* 2004;53(1):109–22.
24. Luximon A, Goonetilleke RS, Tsui KL. [Foot landmarking for footwear customization](#). *ERGONOMICS* 2003;46(4):364–83. [\[PubMed\]](#)

Additional information and reprint requests:

Robert B. Kennedy  
Forensic Identification Research Services  
Royal Canadian Mounted Police  
1200 Vanier Parkway, NPS Bldg.  
Ottawa, ON K1A 0R2 Canada