Science and Justice 52 (2012) 243–248

Contents lists available at SciVerse ScienceDirect

Science and Justice

journal homepage: www.elsevier.com/locate/scijus

The interpretation of shoeprint comparison class correspondences

Sheida Hancock [a,b], Rian Morgan-Smith [a], John Buckleton [a,*]

[a] ESR, Private Bag 92021, Auckland, New Zealand
[b] University of Auckland, Private Bag 92019, Auckland, New Zealand

ARTICLE INFO

Article history:
Received 3 April 2012
Received in revised form 5 June 2012
Accepted 8 June 2012

Keywords:
Forensic
Shoeprint
Interpretation
Class correspondence

ABSTRACT

The underlying principles involved in the interpretation of shoeprint comparisons have become a topical subject due to criticisms in the 2009 National Academy of Science (NAS) report on forensic sciences [1]. Difficulties in the application and understanding of these principles were also highlighted in a recent court ruling [2–5]. We report here a survey that may inform some aspects of this interpretation and discuss the implications of findings from this survey in the light of that court ruling and more importantly the NAS report. Five hundred shoeprints taken from student volunteers in Auckland, New Zealand were compared against each other for the presence of any pattern correspondences. Comparisons were undertaken of the full outsole and of smaller portions of the more common patterns.

Of the 500 shoe impressions collected 488 (97.6%) were ultimately represented only once in the survey. The greatest number of corresponding patterns was for the most common brand of shoe (Converse Chuck Taylor All Star) and occurred in 3 of 500 observations. No instances of an imitation brand matching the authentic brand were found. Smaller sections of the common patterns showed a greater number of corresponding prints. However, the greatest number of matching partial patterns was again for the most common brand of shoe (Converse Chuck Taylor All Star) and occurred in 29 of 500 observations.

We conclude that pattern match alone is of considerable evidential value even when the print is partial.

1. Introduction

In forensic footwear comparison a shoeprint associated with a crime may be compared with test impressions made by a shoe from a person of interest. Prints collected from crime scenes for comparisons are often referred to as crime prints or scene prints. Prints collected from shoes of interest (suspect shoes) for comparison are often referred to as suspect prints or reference prints.

The comparison of scene prints with reference prints starts with an exclusionary process. If there are inexplicable differences between the scene and the reference prints then the shoe is deemed to be excluded as having made the scene print.

If there is adequate clarity in the scene print so that sufficient individualising detail, such as nicks and cuts, may be seen to correspond between the scene and reference prints then the examiner would typically declare that the shoe made the scene print with certainty. This conclusion is termed as either an identification or an individualisation.

There is a valid discussion at this time about whether “identification” is a conclusion that may be scientifically sustained [1–4]. This discussion is oblique to the matter we report here; therefore, we avoid discussing the subject further.

However, it is commonly the case that the scene print, whether it represents an entire shoe sole or just a portion of a sole, is of a reduced clarity or size and little or none of the finer individualising detail present on the shoe sole is visible.

In practice, for the vast majority of cases, the level of correspondence falls somewhere in the spectrum between exclusion and identification.

A situation often occurs when there is sufficient clarity to determine that the overall sole pattern corresponds but any potentially individualising detail is not present or cannot be relied upon because of other interferring background features.

The correspondence of the generic manufacturing features alone is termed a class match [5]. It is the assessment of the evidential value of these class matches that we seek to advance with this work. In determining the evidential value of a class match we produce a baseline from which further assessment of any wear and damage features present can be made.

2. Background

Differing styles of scale have been offered to deal with reporting conclusions which fall between exclusion and identification. Three examples of scales for the reporting of shoeprint comparison are given in Table 1 below. These scales fall into two general styles. The left hand column represents the product of a Bayesian approach to

* Corresponding author.
E-mail address: john.buckleton@esr.cri.nz (J. Buckleton).

1355-0306/$ – see front matter © 2012 Forensic Science Society. Published by Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.scijus.2012.06.002

interpretation. The central and right hand column report posterior probabilities.

The ENSFI marks working group offers two scales; hence two columns have the same label. We have appended the labels ENSFI 1 and 2 to these columns. Many laboratories report all conclusions below inconclusive as an exclusion.

Although the table gives the impression of an alignment between the two constructs no such alignment is possible except at the extreme ends of the scale. The Bodziak and ENSFI 2 scale on the right hand side of the table have an intuitive appeal, an appeal shared by the Court of Appeal in a recent ruling, R v T, in the UK [8–11]. However, the levels on these scales are examples of posterior probabilities and cannot be supported by the scientific examination alone. Such posterior probabilities require the consideration of non-scientific evidence. It is not our intention to reopen this discussion here as it has been dealt with in other publications [3,12–20].

If we come finally to the subject of class match evidence, how is such evidence to be assessed? This is typically done in either a Bayesian or a frequentist construct as exemplified by the two styles of reporting discussed above.

Evett and Weir gave a number of principles of evidence interpretation (see Evett and Weir [21] pg. 29). One of these is that evidence cannot be assessed by considering only one hypothesis. The probability of the evidence, E, must be assessed under each of two competing hypotheses. In the shoeprint context this means that either the suspect shoe made the scene print, Hp, or another shoe made this print, Hd. The ratio of the probability of the scene print under each of these two hypotheses is termed the likelihood ratio (LR). This is the philosophy and essence of the Bayesian construct, and is the recommended way in which shoe impression, and indeed all, evidence should be interpreted. Symbolically this may be written as

$$\mathrm{LR} = \frac{\Pr(E|\mathrm{Hp})}{\Pr(E|\mathrm{Hd})}$$

If we assume that the class features must correspond if the reference shoe made the scene print then $\Pr(E|\mathrm{Hp}) = 1$. $\Pr(E|\mathrm{Hd})$ represents the probability of the scene impression if it was made by a shoe other than the suspect shoe [22].

In the case where the numerator of the LR is 1 the Bayesian and frequentist approaches converge, at least numerically, if not philosophically. The frequentist construct simply reports $\Pr(E|\mathrm{Hd})$ or describes the match in such a way that the fact finders might make some assessment of this probability.

In R v T, the Court of Appeal expressed concerns about the size and relevance of the survey that informs $\Pr(E|\mathrm{Hd})$. Since it is not our intent to relitigate the R v T case here we simply use these matters to highlight issues that commentators may be concerned about.

In expressing $\Pr(E|\mathrm{Hd})$ it is vital that serious consideration be given to the survey's relevance. In fact we would emphasise relevance over the size of a survey.

Therefore, in the sections below those matters relating to the relevance of surveys of shoeprints will be discussed. However, in the course of actual casework most shoeprint examiners have become aware that clarity and completeness of the shoe print are the most crucial factors in determining the outcome of any comparison. We will also discuss the impact of these factors.

2.1. Relevance of survey data

At this point we could conceptualise the perfect survey to inform $\Pr(E|\mathrm{Hd})$. Considering this probability we need to think how we could inform it considering that the print was detected at a certain location and at a certain time but was not made by the shoe of the person of interest (recall that we are assessing this probability under Hd). A survey of shoeprints at exactly similar locations and at exactly this date is the best we can envisage, but is clearly impractical.

It is advisable for scene examiners to attempt to ensure that the shoeprints that they take have a reasonable chance of being deposited by the offender as opposed to being random background shoeprints left by persons innocently visiting the premises. In doing this they may consider such factors as:

1. Location of the print, is it near the point of entry or in a much travelled area,
2. Elimination prints from homeowners and first responders, and
3. Appearance of the print with respect to age.

However, undoubtedly some shoeprints from the natural background of the environment will also be collected. This would affect our thinking. If the shoeprint is definitely left by an offender then we need a survey that models potential offenders. If the shoeprint may have been left by anyone of the general public we need to consider a survey of a more general nature.

We seek to assign a value to the probability $\Pr(E|\mathrm{Hd})$. There is a worrying misconception that there is such a thing as the right number for such a probability. We need to consider what survey or surveys may best inform our assessment of this probability. In this there is an unavoidable subjective element. Most obviously in this context this is represented by our assessment of what survey may be the most appropriate [23].

If we have some confidence that the shoeprint was left by the offender then a reasonable approximation might be a survey of offenders' shoes in the relevant area and within a reasonable time of the

Table 1
Three of the published shoeprint conclusion scales showing the two primary styles.

Part of the ENSFI marks working group conclusion scale [6,7]. Termed ENSFI 1.	Bodziak [5] pg. 372–374	Part of the ENSFI marks working group conclusion scale [6,7]. Termed ENSFI 2.
Identification	Positive identification	Identification
Very strong support for proposition A	Probably made	Very probably
Strong support		
Moderately strong support		Probably
Moderate support	Possibly made	
Limited support		
Inconclusive		Inconclusive
Limited support for proposition B	Possibly did not make	Likely not
Moderate support		
Moderately strong support		
Strong support		
Very strong support		
Elimination	Non-identification	Elimination

offence. Since offenders seldom identify themselves we need to, again, approximate the population of offenders. The Forensic Science Service (FSS) survey of suspects' footwear within the 3 years preceding the offence used in R v T does seem to be a reasonable compromise.

A survey of all footwear sales is likely to have significant contributions from footwear styles which we believe would not reflect the footwear choices of the typical offender. Sales of sports shoes might be nearer the required survey since most footwear impressions submitted in casework are from sports shoes. Sales figures, however, do not give the pattern of the outsole, may not give the geographical distribution, extent of losses, defects in distributed shoes, or extent of rebranding.

In R v T there is speculation by the court that imitation brands might approximate the outsole pattern of the authentic brand. In our own jurisdiction we are aware that popular brands often have many moulds for the common sizes and these can be differentiated in clear prints. It is therefore very likely that imitation brands differ in subtle ways from the authentic brand. However, whether or not the imitation brand is a perfect rendition of the outsole pattern copied, the brand of the manufacturer is immaterial to assessing the frequency of the actual outsole pattern.

In New Zealand the volume of casework for which the prints are taken is such that a survey of suspects' shoes would produce very few data.

Conviction data from New Zealand indicate that 17–29 year old offenders are responsible for 60% of homicides and 76% of breaking and entering offences. Approximately 80% of convicted offenders are male [24]. Without meaning any disrespect to tertiary students, a group the authors were all once members of, they may share some attributes with offenders at least as far as age, income and means of transport. Therefore, in this work it seems reasonable to use tertiary students as a surrogate for the hypothetical offender survey.

There is a logical connection from data to statistic to probability to expert evidence. A statistic is any function of the data and as such follows directly from the data. However the transitions from statistic to probability and from probability to expert evidence require judgement. These judgements revolve around, in this case, relevance of the data to the current case. The discussion above, we hope, emphasises that subjective judgement cannot and should not be removed from the process of assigning a probability.

2.2. Size

The size of the survey was a matter discussed at some length in R v T and from the court's ruling they appear disappointed at the FSS survey of 8122 shoes from suspects. They contrast this with the approximately 42 million sports shoes sold a year in the UK. In this emphasis the court was almost certainly astray. The uncertainty in a sample frequency from a survey is not dependent on the fraction of the population surveyed until a significant fraction of the whole population has been surveyed. Rather it is dependent on the number actually sampled.

Sampling uncertainty in surveys is a well studied field. The uncertainty in the sample frequency from a survey may be estimated in a number of well investigated ways from standard statistical theory. If the estimate itself is not small then a confidence interval based on the normal approximation will be adequate.

However, there are so many different outsole patterns that each one tends to be infrequent. Even the most common brands appear to show variations in mould design, often even for a single common size and hence even for the common brands each outsole pattern may be infrequent. For such situations a confidence interval based on the normal approximation is likely to be inadequate and the exact confidence interval suggested by Clopper and Pearson (C&P) [25] or a probability interval based on the region of highest posterior density (HPD) would be preferable. In this aspect there is a strong parallel between sampling uncertainty in shoeprint work and in the interpretation of lineage markers in DNA [26]. In the appendix we outline the detailed methodology for the C&P confidence interval and also HPD. However, the reader may be interested that the sampling uncertainty for an estimate of 20% in a sample of 8122 as given by Mr Ryder in the R v T case suggests a 95% probability interval of 19.1–20.9%. In other words there is a 0.95 probability that the probability in the population from which this data is obtained is between 19.1 and 20.9%. This would suggest that sampling uncertainty in a survey of 8122 was unlikely to be a significant factor that should have affected the court's decision making.

The authors advocate that the uncertainty in estimates arising from sample size should be calculated and plausibly presented along with the estimate. However, we repeat, that the relevance of survey data may be, and often is, of greater importance than the size of the survey.

2.3. Clarity and completeness

Shoeprints at scenes vary greatly in clarity and completeness. The more that may be seen of a shoeprint with good clarity the more likely that subtle mould differences and individualising details may be differentiated. Therefore, the probability $\Pr(E|\mathrm{Hd})$ is determined to some considerable extent by the clarity and completeness of the scene print. In fact clarity may be much more important than completeness. A small but clear partial print is likely to be of more value than a larger but smeared or indistinct print.

The more complete the print and greater the clarity of the detail the more likely the print is a true and accurate representation of the shoe sole. Correspondingly the better the information the analyst has when doing the comparison, the more likely a conclusion will be made that tends to the extremes of the conclusion scales. The smaller the print or less distinct the detail the more likely the conclusion will be based on the gross manufacturing detail. Even with only gross manufacturing or class features visible an exclusion is often readily evident. It is those cases where the scene print and reference print share class detail but there is insufficient individualising detail that we refer to as a class match. For these cases, as long as the correspondence of the wear and detail that can be seen is good, then $\mathrm{LR} \approx \frac{1}{\Pr(E|\mathrm{Hd})}$). It is the probability $\Pr(E|\mathrm{Hd})$ that we seek to inform in this survey.

3. Method

In March 2010, 500 footwear impressions were collected from the right shoes of people at two University of Auckland, New Zealand, campuses. No sex, age or ethnicity data were collected to comply with our ethics approval. Shoeprints were collected using the Identicator LE-25 inkless shoeprint kit.

A booth was set up in a main public thoroughfare near the main cafeterias. Of the passing foot traffic as many individuals as was practically possible were asked to provide shoeprint impressions. Any consenting participant was sampled. Each print was numbered and the size and model of shoe, if known, was recorded. No other information was collected from the donors.

The prints are described hereafter as full but they were only as full as the participant printed. In some cases regions of the sole pattern, often at the toe, were missing.

Each print was compared with each other resulting in 124,750 comparisons, where it was known that the shoes are different. This method has been termed Tippett testing [27]. It was refined and used extensively by Evett (see for example Ref. [21]) and has subsequently been used by many others in examining the performance of forensic interpretation methods [28–30].

The decision process reported here was made at two different levels of discrimination for prints showing the full outsole.

1. Grouping based on generic pattern match only and ignoring the effects of wear, size, and details of the configuration of mould features, and
2. Further differentiation within the groups with consideration given to details of the sole pattern such as the relative placement of mould elements.

The first grouping approach simplified the subsequent comparison work.

In casework a full scene print of the outsole at perfect clarity is rare. It is much more common to be dealing with partial prints, often with small areas of clear detail.

To mimic these types of partial prints 5 of the 500 survey prints were selected using a random number generator. This set was augmented by two prints from the most abundant pattern grouping (Converse Chuck Taylor All Star pattern, CCTAS) and one from the second most abundant pattern group (Kmart casual canvas shoe pattern, KCC). This gave a total of eight selected outsole impressions. Two partial prints were created from each of these eight shoes using a circular stencil 50 mm in diameter. A section was selected from the ball area and from the heel area, being two of the more common contact areas for the creation of shoeprints. These 16 partial prints were compared blind with the 500 surveyed prints. Each of the 16 prints therefore had its true parent in the test set plus 499 potential false matches.

For the last experiment one shoe was selected to perform a 25 mm diameter partial profile test. This shoe was of the most common brand (CCTAS). It was not one of the shoes from within the survey and could be differentiated on class features from each of the survey prints. The sole pattern of the CCTAS is comprised from three main pattern features. Nine 25 mm diameter areas were selected from the outsole of which three showed one sole pattern feature, three showed two sole pattern features and a further three showed three sole pattern features. These were then compared blind with the 500 surveyed prints.

In some cases a comparison was not possible because the print being compared against had not printed in the area being examined, often the toe. Hence the sample size, in some cases, is fewer than 500.

4. Results

Results of the grouping process based on considering pattern match only and ignoring the effects of wear, size, and details of the configuration of mould features are given in Table 2.

The distribution of pattern groups as shown in Table 2 shows many low frequency groups and a few high frequency ones. The presence of high and low frequency groups is also seen in a survey from Ireland by Hannigan et al. [31]. The pattern of some high and many low frequencies patterns could reasonably be expected if we believe that some manufacturers or brands of shoe will have a larger market share by volume.

In Table 2 the group with 55 members was composed of prints from CCTAS shoes. This is a very popular shoe which has been in production for many years with very little change to its outsole pattern. The group with 21 members was composed of prints from a casual canvas shoe. This was a very cheap shoe with over 10 different types or style of canvas uppers that share the same broad outsole features and was sold at Kmart.

Table 2
Grouping based on generic pattern match.

Generic pattern group size	Number of groups	Generic pattern group size	Number of groups
1	276	6	1
2	28	7	1
3	10	11	2
4	3	21	1
5	3	55	1

Table 3
Size of class match groupings.

Brand of shoe	Size of pattern group	Number of class subgroups	Size of subgroups (>1)
CCTAS	55	50	3, 2, 2, 2
Vans canvas Era	11	8	3,2
Nike Air Force 1	11	9	2,2
Dr. Martens 1461	6	5	2
Winsor Smith Flashy	5	4	2

Each of the prints within the pattern grouping was then compared with respect to the specific detail of outsole tread features size, location and orientation to identify actual class matches in sole patterns. In Table 3 we give the groups remaining after this detailed comparison. Five of the pattern groups still had shoes that could not be distinguished based on actual class features.

The CCTAS group had 46 sole patterns that could be distinguished from all others. There were three groups where two outsole patterns could not be distinguished from each other and one group where three outsole patterns could not be distinguished from each other. All of the KCC shoes could be distinguished from each other.

Of the 500 shoe impressions collected 488 (97.6%) were ultimately represented only once in the survey.

The results of the comparisons for the 16 50 mm partial prints are summarised in Table 4. Each of the randomly generated partial prints only provided a class match to their parent print and did not provide a class match with any other print within the survey.

The two selected KCC partials and two of the selected CCTAS prints only provided a class match to their parent print and did not provide a class match with any other print within the survey. The CCTAS partial print 1 provided a class match to 29 other prints, including its parent print. However, this partial print had only one pattern feature visible which was the characteristic diamond pattern. In addition the region selected showed significant wear which obscured some of its features. If the wear features were taken into consideration during the comparison then only seven of the other CCTAS prints provided an actual class match. The CCTAS partial print 2 was from the heel of the shoe and displayed multiple pattern features yet it still provided a class match to three other prints, including its parent print.

The results of the 25 mm partial CCTAS print experiment are given in Table 5. In some cases a comparison was not possible because the print being compared against had not printed in the area being examined, often the toe. Hence the sample size, in some cases, is fewer than 500.

These smaller CCTAS partial prints could still all be differentiated from any non CCTAS shoes in the survey.

These partial prints did provide multiple class correspondences particularly when only one sole pattern feature was available for comparison. For one partial print there was a class match to twenty-three other CCTAS shoes in the survey. However this particular print only displayed the parallel line pattern in the arch of the outsole pattern.

Table 4
Results of the 50 mm partial print comparisons.

Partial print	Result of comparison
Partial prints from the heel or toe of the 10 randomly selected outsoles	Match to parent only
CCTAS; partial print 1 (ball)	29/500 (7/500)
CCTAS; partial print 2 (heel)	3/500
CCTAS; partial prints 3 (ball) & 4 (heel)	Match to parent only
KCC partial prints 1 & 2	Match to parent only

Table 5
Results of the 25 mm partial print comparison.

Number of pattern features	Results of comparison matches/revised sample size		
	Area 1	Area 2	Area 3
1	12/482	13/500	23/500
2	11/491	7/500	5/500
3	5/490	7/500	2/500

5. Discussion

The general finding of this work is that in March 2010 in Auckland, New Zealand, all shoeprints are rare when at full clarity and completeness. Even the most common brands show considerable variation within the sole pattern arising from mould variations. No instance of an imitation brand corresponding with the authentic brand was found. This has been documented previously and is expected [5].

In Fig. 1 below we give the distribution of class matching pattern group size for the survey reported here and the one reported by Hannigan et al. [31]. We note a distinct similarity in the distribution of common and rare patterns. There is data at group sizes over 10, but at very low frequency.

The vast majority of outsole patterns occur very rarely but there are a very small number of outsole patterns, or brands of shoe, which are over represented. These over-represented shoes are likely to be brands that are well known and acknowledged to be popular and are likely to have been popular over considerable periods of time with little or no variation to their construction.

Even when the shoeprint was partial very clear discrimination still remained. With the 50 mm partial prints the only class matches to non parent prints were from the most commonly occurring shoe being the CCTAS.

The comparison of the smaller 25 mm CCTAS partial prints still provided some discrimination.

It may be that the limited number of outsole features and their limited configurations which combine to form the uniquely identifiable CCTAS outsole provide a worst case scenario for this type of comparison.

The different styles and scales in the reporting of shoeprint conclusions highlight the absence of agreement and conformity. The Bayesian reporting practices of the authors' organisation have provided verbal descriptors for the orders of likelihood ratio and although the numerical application has been primarily for DNA analysis the verbal scale has been applied to the range of shoeprint comparison conclusions. This is given in Table 6. It will be noted that this scale has broad similarities with the Bayesian framework of the ENSFI 1 scale.

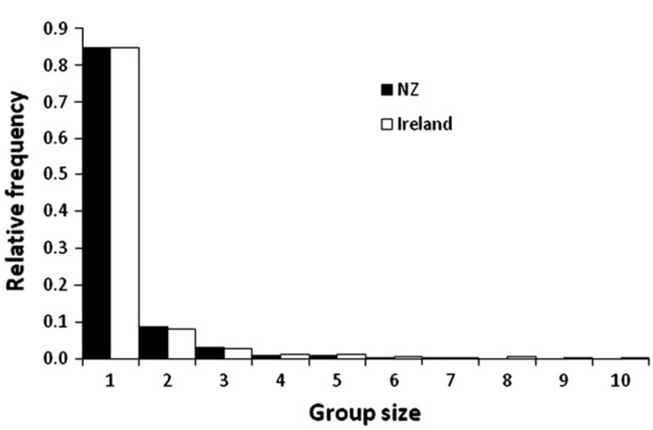

Fig. 1. The relative frequency for group sizes 1–10 for the NZ and Irish surveys.

Table 6
The verbal equivalents used at ESR, New Zealand.

Likelihood ratio	Verbal equivalent
1	Is neutral
1–10	Provides slight support
10–100	Provides moderate support
100–1000	Provides strong support
1000–1,000,000	Provides very strong support
Over 1,000,000	Provides extremely strong support

Data obtained from this work have allowed likelihood ratios to be assigned which provides a numerical basis for categorising a class match.

6. Conclusion

We expected there to have been very considerable variation in outsole patterns. With the exception of outsole patterns acknowledged to be popular, or common, any outsole pattern is likely to be rare and to represent less than 0.2% of the population. This suggests that a class match undertaken in isolation to other more detailed comparisons, would still be of considerable evidential value.

In our work the most common outsole pattern occurred in 3 of 500 observations for a full print suggesting an LR of over 100 with a lower bound of the 99% probability interval of 60. This suggests that, with probability 0.99, the LR is greater than 60.

The least frequent patterns clearly are those that we have not observed at all. The lower bound of the 99% probability interval for an LR for these would be 130.

An assessment of whether or not a particular shoeprint, and its corresponding shoe, is common or rare must be a crucial consideration in performing a shoeprint comparison.

In a case where a class match is obtained between the reference print and a full scene print, for a sole pattern known to be a commonly occurring pattern then, the evidence is estimated to be at least 60 times more likely if it comes from this shoe than if it was made by another shoe. This would provide moderate support for the suggestion that the suspect shoe made the scene print.

In a case where a class match is obtained between the reference print and a full scene print, for a sole pattern known to be a rarely occurring pattern then, the evidence is estimated to be at least 130 times more likely if it comes from this shoe than if it was made by another shoe. This would provide strong support for the suggestion that the suspect shoe made the scene print.

Clear but small partial prints lose some, but certainly not all, of the evidential value of the parent outsole. The lower 99% probability interval for the *LR* for the 50 mm CCTAS partial prints ranged between 14 for one outsole feature and 77 for three outsole features. This suggests that a class match for a small partial print, even of a commonly occurring style, would still provide moderate support for the suggestion that the suspect shoe made the scene print.

Acknowledgements

We warmly thank Sally Coulson, Dion Sheppard and two anonymous referees whose comments greatly improved this paper.

Appendix 1. Sampling uncertainty

The count, x, of a certain sole pattern, from a database of size N, is expected to follow a binomial distribution. The binomial estimate, $\tilde{p} = \frac{x}{N}$, is an unbiased estimate of the population proportion. A confidence interval based on a normal approximation to the binomial $\tilde{p} \pm z_{\frac{1-\alpha}{2}}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{N}}$ may be inappropriate for low $\tilde{p}$ because of the right hand skew to the distribution.

Both frequentist and Bayesian solutions to the problem exist.

A.1. Frequentist exact confidence interval

A frequentist solution was suggested by Clopper and Pearson [25] which seeks a value p_0 such that $\sum_{k=0}^{x}\binom{N}{k}p_0^k(1-p_0)^{N-k}=0.05$. This formula may be implemented easily in EXCEL™. An EXCEL™ layout is given at the end of this appendix.

A.2. Bayesian highest posterior density

The likelihood of observing x successes out of N trials is given by the binomial distribution $\binom{N}{x}p_i^x(1-p_i)^{N-x}$. To form a posterior we require a prior. If the sampling is binomial $\mathrm{Bin}(x,N)$ and the prior is beta $\beta[\alpha,\beta]$ the posterior is also beta $\beta[x+\alpha,N-x+\beta]$. This distribution is readily implemented in EXCEL™ and the quantiles from the posterior distribution may be obtained.

A uniform prior $\beta[1, 1]$ places equal probability density on all values of p_i between 0 and 1 and in the case of shoeprints is highly conservative. Whichever prior is chosen it is combined with the likelihood to give the posterior probability.

Since courts, if they are interested in sampling uncertainty at all, are usually interested in the upper bound we advocate a one sided interval.

An EXCEL™ layout for implementing the frequentist exact confidence interval and the HPD for a $\beta(1,1)$ or any prior.

	A	B	C	D
1	x	3	α	1
2	N	500	β	1
3				
4	C&P 95% one sided confidence interval =VLOOKUP(0.95,$A$7:$C$1007,3)	95% one sided probability interval =VLOOKUP(0.95,$B$7:$C$1007,2)		
5	0.0154	0.0154		
6	C&P	HPD(1,1)	P_0	
7	=1-BINOMDIST($A$1,$A$1,C7,1)	=BETADIST(C7,$B$1+$D$1,$C$2-$C$1+$D$2)	0.0000	
8	0.0000	0.000	0.0002	
9	0.0002	0.000	0.0004	
10	0.0004	0.000	0.0006	
11	0.0006	0.000	0.0008	
81	0.938	0.939	0.0148	
82	0.942	0.943	0.0150	
83	0.946	0.946	0.0152	
84	0.949	0.950	0.0154	
85	0.953	0.953	0.0156	
86	0.956	0.956	0.0158	
87	0.959	0.959	0.0160	
88	0.962	0.962	0.0162	
89	0.964	0.964	0.0164	

References

[1] D.A. Stoney, What made us ever think we could individualize using statistics, Journal of the Forensic Science Society 31 (2) (1991) 197–199.
[2] M.J. Saks, J.J. Koehler, The individualisation fallacy in forensic science evidence, Vanderbilt Law Review 61 (1) (2008) 199–219.
[3] S.A. Cole, Forensics without uniqueness, conclusions without individualisation: the new epistemology of forensic identification, Law, Probability and Risk 8 (2009) 233–255.
[4] S.A. Cole, Where the rubber meets the road: thinking about expert evidence as expert testimony, Villanova Law Review 52 (2008) 101–115.
[5] W.J. Bodziak, Footwear impression evidence—detection, recovery and examination, in: V.J. Geberth (Ed.), Practical Aspects of Criminal and Forensic Investigations, 2nd edition, CRC Press, Boca Raton, 2000.
[6] G. Volckeryck, Fifth meeting of the committee on the harmonisation of conclusion scales in footwear casework, The Information Bulletin for Shoeprint/Toolmark Examiners 8 (1) (2002) 17–32.
[7] Conclusion Scale Committee, Conclusion scale for interpreting findings in proficiencty tests and collaborative exercises within the WG marks, The Information Bulletin for Shoeprint/Toolmark Examiners 11 (2) (2005) 15–18.
[8] R v T, Neutral Citation Number: [2010] EWCA Crim 2439, Court of Appeal, 2010.
[9] C.E.H. Berger, et al., Evidence evaluation: a response to the court of appeal judgment in R v T, Science & Justice 51 (2) (2011) 43–49.
[10] O.E. Facey, R.J. Davis, Re: Expressing evaluative opinions; a position statement, Science & Justice 51 (2011) 212.
[11] Expressing evaluative opinions: a position statement, Science & Justice 51 (1) (2011) 1–2.
[12] H. Katterwe, True of false? The Information Bulletin for Shoeprint/Toolmark Examiners 9 (2) (2003) 18–25.
[13] F. Taroni, J. Buckleton, Comments of Horst Katterwe to the article of F. Taroni and J. Buckleton, The Information Bulletin for Shoeprint/Toolmark Examiners 8 (3) (2002) 17–21.
[14] C. Champod, et al., Comments on the scale of conclusions proposed by the ad hoc committee of the ENFSI Marks Working Group, The Information Bulletin for Shoeprint/Toolmark Examiners 6 (3) (2000) 11–18.
[15] C. Champod, G. Jackson, Comments on the current debate on the Bayesian approach in marks examinations, The Information Bulletin for Shoeprint/Toolmark Examiners 8 (3) (2002) 22–25.
[16] FSS, Reproaches of Forensic Science Service, The Information Bulletin for Shoeprint/Toolmark Examiners 8 (1) (2002) 23–24.
[17] C. Champod, F. Taroni, P. Margot, Statistics: a future in tool mark comparison? AFTE Journal 28 (4) (1996) 222–229.
[18] H. Katterwe, Comments/objections to reproaches of Forensic Science Service and University of Luasanne, The Information Bulletin for Shoeprint/Toolmark Examiners 8 (1) (2002) 25–30.
[19] I. Keereweer, Comments and objections to reproaches, The Information Bulletin for Shoeprint/Toolmark Examiners 8 (1) (2002) 31–32.
[20] A. Biedermann, F. Taroni, P. Garbolino, Equal prior probabilities: can one do any better? Forensic Science International 172 (2–3) (2007) 85–93.
[21] I.W. Evett, B.S. Weir, Interpreting DNA Evidence—Statistical Genetics for Forensic Scientists, Sinauer Associates, Inc., Sunderland, 1998.
[22] I.W. Evett, J.A. Lambert, J.S. Buckleton, A Bayesian approach to interpreting footwear marks in forensic casework, Science & Justice 38 (4) (1998) 241–247.
[23] C. Champod, I.W. Evett, G. Jackson, Establishing the most appropriate databases for addressing source level propositions, Science & Justice 44 (3) (2004) 153–164.
[24] Statistics New Zealand, Convicted offenders by ASOC, 2009.
[25] C.J. Clopper, E.S. Pearson, The use of confidence or fiducial intervals illustrated in the case of the binomial, Biometrika 26 (1934) 404–413.
[26] J.S. Buckleton, M. Krawczak, B.S. Weir, The interpretation of lineage markers in forensic DNA testing, Forensic Science International: Genetics 5 (2) (2011) 78–83.
[27] C.F. Tippett, et al., Paint flakes from sources other than vehicles, Journal of the Forensic Science Society 8 (2) (1965) 61–65.
[28] P. Gill, et al., Interpretation of complex DNA profiles using empirical models and a method to measure their robustness, Forensic Science International: Genetics 2 (2) (2008) 91–103.
[29] P. Gill, J. Curran, C. Neumann, Interpretation of complex DNA profiles using Tippett plots, Forensic Science International: Genetics Supplement Series 1 (1) (2008) 646–648.
[30] G.S. Morrison, Measuring the validity and reliability of forensic likelihood-ratio systems, Science & Justice 51 (2011) 91–98.
[31] T.J. Hannigan, et al., Survey of 1276 shoeprint impressions and development of an automatic shoeprint pattern matching facility, Science & Justice 46 (2) (2006) 79–89.