

Use of data to inform expert evaluative opinion in the comparison of hand images—the importance of scars

Graham Jackson · Sue Black

Received: 4 November 2012 / Accepted: 24 January 2013 / Published online: 5 February 2013
© The Author(s) 2013. This article is published with open access at Springerlink.com

Abstract Evaluation of a likelihood ratio is widely recognised as the most logical and appropriate means of assessing and expressing the weight of expert scientific evidence. This paper describes the application of such an approach to cases involving the comparison of images of hands that contain visible scars. Such evidence is frequently provided in cases of alleged child sexual abuse in which images of the perpetrator's hand are compared with images of the suspect/accused's hand. We illustrate how data provided from a database of hand images can be used to inform the probabilities that are an essential part of evaluating a likelihood ratio and, hence, how data have a bearing on the appraisal of the weight of evidence that can be attributed when scars are present within an image.

Keywords Likelihood ratio · Scars · Hand · Forensic · Identification

Introduction

In 2009, the Association of Forensic Science Providers (AFSP) published a standard for the provision of expert evaluative opinion [1]. The standard was based on extensive, existing scholarship on the logically correct way of appraising the weight of evidence [2–5] and on the nature of expert opinion [6, 7], rooted in a Bayesian paradigm of inference.

Two broad types of expert opinion have been proposed—‘investigative’ and ‘evaluative’. As the name implies, investigative opinions are provided in answer to questions that

are relevant, generally but not exclusively, to the investigative stages of an enquiry, before a suspect is apprehended. Questions of importance in this phase of a case may relate to the type of activity that has taken place and, in a subsidiary role, the potential sources of ‘questioned’ material recovered from the scene or the victim. In contrast, evaluative opinions are provided to help the courts of law to answer questions that are of importance to them. Note that, in this mode, the expert is not answering the question; the expert is providing an expert evaluation of the weight of the observations from the tests or procedures in order to help the court arrive at an answer. Generally, evaluative questions will relate to whether the accused has indeed carried out a particular activity that forms a component part of the offence. Alternatively, in some cases, the questions may relate to subsidiary issues including whether questioned material recovered from the scene or from the victim belongs to the accused.

While there may well be investigative questions that can be addressed through an examination of hand images, we will limit this communication to a consideration of evaluative questions. Through the use of three case studies, we will illustrate how the AFSP standards [1] may be applied to such cases, particularly focussing on the use of data to inform probabilities. In so doing, we will show how evaluative opinion may be formed and communicated.

Interpretative framework

The AFSP standards [1] describe how an expert operating in ‘evaluative’ mode should offer an opinion in the form of a likelihood ratio (LR). For those readers unfamiliar with the concept of a likelihood ratio, we include here a brief explanation of the LR within the framework of Bayes’ theorem. We direct readers to other sources for further details [2–7].

G. Jackson
Advance Forensic Science, University of Abertay, Dundee, UK

S. Black (✉)
Centre for Anatomy and Human Identification, University of Dundee, Dundee, UK
e-mail: s.m.black@dundee.ac.uk

Bayes' theorem is a formalisation of inductive inference, demonstrating how newly acquired information may be used to update, in a logical way, the probability of an uncertain event. In the context of forensic science, Bayes' theorem provides the logically coherent means by which the 'triers of facts', on being presented with the expert's observations, can move forward their view on the probability that an allegation was in fact true. One essential element of Bayes' theorem, as used in judicial settings, is an appraisal of the weight of the evidence, i.e. the impact that the expert's observations should have on the tribunal's opinion on the truth of the allegation. The weight of evidence is provided by the log of a value known as the LR. In simple terms, the LR is the ratio of the probability of the expert's observations, given a proposition (allegation) were true, to the probability of obtaining the same observations, given an alternative proposition were true. The odds of the proposition being true, prior to the expert's observations, are multiplied by the LR to form the new a posteriori odds of the proposition being true.

It is useful to capture the concept of the likelihood ratio in mathematical terms and, in situations in which there is a clear prosecution proposition and an alternative proposition for the defence, the LR can be represented by:

$$LR = \frac{\Pr[E|H_P, I]}{\Pr[E|H_D, I]} \quad (1)$$

where:

$\Pr[E | H_P, I]$ represents the probability (Pr) of obtaining the observations (E) given the truth of the prosecution proposition (H_P) and given the relevant case circumstances (I);

$\Pr[E | H_D, I]$ represents the probability (Pr) of obtaining the observations (E) given the truth of the defence proposition (H_D) and given relevant case circumstances (I).

The expert's role in this mode is to use their expert knowledge and whatever relevant datasets there may be to assign probabilities (Pr) for the observations (E) given that, respectively, the prosecution (H_P) and defence propositions (H_D) were true, and given the conditioning aspects (I) of the case circumstances. The weight of evidence can then be expressed as the log of the likelihood ratio or as a verbal scaled equivalent of that value.

Propositions flow from the issues that are identified as being of interest and importance to the triers of fact. Generally, in cases involving a comparison of hand images, the issue will be—'is the hand in the perpetrator's image the hand of the accused'? The prosecution proposition that flows from that issue would be:

H_P —The image is that of the hand of the accused.

If the accused denies that the image is of their hand and offer no alternative explanation, then the defence's alternative proposition would most likely be:

H_D —The image is that of some other, unknown, person.

However, the precise specification of the alternative proposition would be informed by aspects of the case circumstances (I) that give information on the offender. Other aspects of the case circumstances (I) would have a bearing on the probabilities for the observations, and for hand comparison purposes, these might include:

1. The time interval between the offender's image being recorded and the reference images being taken from the suspect/accused
2. The recording conditions (lighting and resolution) of both sets of images

A refinement of the evaluation of an LR, relating specifically to the assignment of conditional probabilities for the observations (E), has been proposed by Champod et al. [8]. The authors offer guidance on this task by breaking down the observations (E) into two component parts and by making some simplifying assumptions. Following the scheme and notation as presented by these authors and adopting two of their three assumptions, the LR in cases involving comparison of hand images can be simplified to:

$$LR = \frac{\Pr[O_C|O_D, H_P, F_C, F_D]}{\Pr[O_C|H_D, F_C]} \quad (2)$$

where:

O_C denotes the observations from the perpetrator's image.

O_D denotes observations from the suspect/accused's image.

F_C represents the circumstances surrounding the incident, including evidence about the nature of the true perpetrator and the recording conditions of the questioned images.

F_D represents the circumstances surrounding the suspect/accused, including the time interval between the incident and the suspect/accused's images being taken and the recording conditions of the reference images.

While the formula may, at first sight, seem daunting, it may be helpful to some readers to express the formula in words.

The numerator asks: what is the probability that the perpetrator's images would have appeared as they do (O_C), given what has been seen in the images of the accused's hand (O_D), given the images are truly of the accused's hand (H_P) and given the relevant circumstances of the crime and the accused (F_C and F_D)?

The denominator asks: what is the probability that the perpetrator's images would have appeared as they do (O_C), given the images are of the hand of some unknown person (H_D) and given the relevant circumstances of the crime (F_C)?

In the formulation given by Champod et al. [8], the numerator of Eq. 2 has been set to a value of 1, i.e. it is certain that the perpetrator's images would have appeared as they do (O_C), given what has been seen in the images of the

accused's hand (O_D), given the images are truly of the accused's hand (H_P) and given the relevant circumstances of the crime and the accused (F_C and F_D)? However, it should be noted that, depending on the time lapse between the two sets of images being taken and the difference in quality between them, the value of the numerator may in reality be somewhat less than 1.

Prior to the AFSP standards being published, Evett et al. [9] described three principles of logical interpretation that flow from accepting a likelihood ratio approach to assess evidential weight. Somewhat paraphrased, these are:

1. Scientific evidence is to be interpreted within a framework provided by the relevant case circumstances.
2. Evidence can only be evaluated where there are at least two propositions, usually reflecting the prosecution and defence positions.
3. Scientists use their expert knowledge and relevant data to assign a probability for their observations, conditioned on the relevant case circumstances and on the propositions that are being addressed to help the triers of fact.

While each of these three principles is worthy of further discussion and consideration in their own right, it is the third principle that seems to raise the most concern among practitioners who are new to this approach. We therefore spend a little time here expanding on this principle.

The notion of assigning a probability for the outcome of a scientific 'test' can seem quite alien both to scientists and to lay people. After all, the outcome of the test is real—it exists. Therefore, the argument could be that the probability of obtaining that outcome equals 1. However, the scientist should ask questions of the form: 'What is the probability that I would have obtained that outcome if a specific (prosecution, defence or other) proposition were true?' Those probabilities will be informed both by the scientist's personal knowledge, based on their own experience, and by whatever relevant data may be available. This method of arriving at a probability for the outcome is embedded in the AFSP standard's [1] requirement for 'robustness' of scientists' opinions. Probabilities in this form are necessarily subjective, although expert, in nature. They are descriptors of a scientist's assessment of the uncertainty of obtaining a particular outcome. There will be no such thing as the one, precise, true number for this probability, even if the probability is based on extensive, well-tested datasets. Datasets do provide precise statistics on, for example, relative frequencies of characteristics within the dataset sample, but the scientist then has to apply expert knowledge and understanding to translate such statistics into subjective, expert probabilities for the specific outcome of the test in the particular circumstances of a unique case [2, 10, 11].

The authors of the present paper have found it helpful, when working with practitioners on the notion of the

subjective nature of probability, to encourage experts to assign probabilities in advance of carrying out the scientific 'test'. This necessarily directs the expert to consider all possible outcomes of the test and their associated probabilities prior to carrying out any substantive work. This process helps to avoid post hoc justification of probability and guards against 'observer bias'. This 'pre-assessment' of the probabilities of outcomes is a vital early step in the process known as 'Case Assessment and Interpretation' [12, 13].

The three principles, and the interpretative model that flows from them, have been presented in the form of a step-by-step aide-memoir that acts as a prompt for the examining scientist to follow during the course of a case. One example of an aide-memoire is provided by Jackson and Jones [13].

This paper illustrates the application of the three principles [9], the AFSP standards [1] and the Champod et al. approach [8] through three case studies concentrating only on information pertaining to scars that can be identified in images of the hand. The case studies are real but anonymised.

Case studies

Case 1. The case of the abandoned Blackberry

Step 1 Relevant case circumstances (F_C and F_D)

Police raided the house of a man suspected of smuggling illegal cigarettes. He ran out of the back door and threw his Blackberry device over a fence. On examination of this device and the man's computers, many indecent images were found. Included in these were images of the hand of a perpetrator performing indecent acts on children. The images had been taken in the house of the suspect. He was charged with various offences, and in response, said the images were not of him. The time interval between the perpetrator's images and the reference images being taken were unknown, but the quality of the perpetrator's images was reasonably good.

Step 2 Case assessment

The images from the perpetrator were examined prior to inspection of the reference images from the accused. From the images of the perpetrator, it could be inferred that he was an adult, light-skinned male. The images were of the left hand only. Two linear scars, one small and one large, were identified at the base of the thumb (Fig. 1). No other scars were apparent in the images. The presence of two scars in this location, and the absence of scars elsewhere, was denoted ' O_C '.

It was noted earlier that the general, default set of propositions in cases such as this would be of the form:

H_P —The image is that of the hand of the accused.

H_D —The image is that of some other, unknown, person.

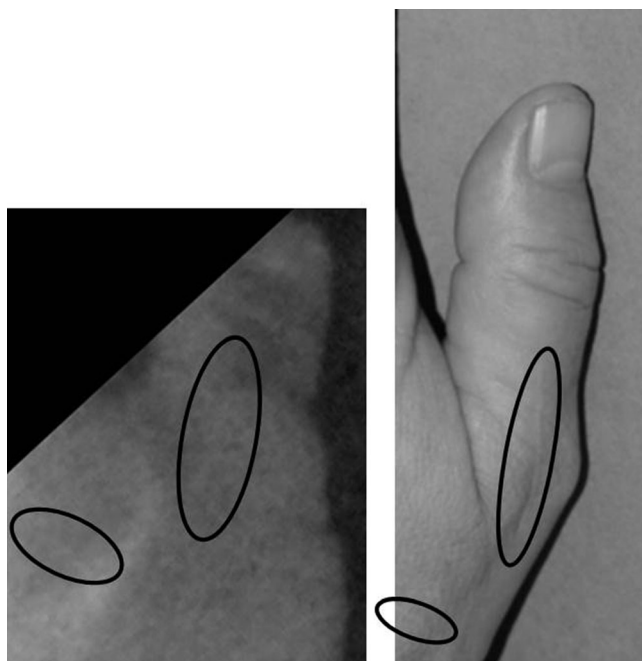


Fig. 1 The perpetrator's image is to the left and that of the accused to the right. Oval outlines show the position of two scars (one large and one small) at the base of the thumb

From the relevant aspects of this specific case, the propositions can be refined as follows:

H_P —The image is that of the left hand of the accused.

H_D —The image is that of the left hand of some other, unknown, light-skinned, adult male.

If, on examination of the accused's images (O_D), the expert were to see the presence of two scars in the same location as in the perpetrator's images (O_C), and no scars elsewhere, the weight of evidence provided by this correspondence could be evaluated following Eq. 2.

Looking firstly at the numerator of this equation, $\Pr [O_C|O_D, H_P, F_C, F_D]$, the expert is required to assess the probability that the perpetrator's images would show the features that they do (O_C), given that they truly are those of the accused's left hand (H_P) and given what is known about the conditions of recording the perpetrator's images (F_C) and about the accused and his images (F_D). This could be considered to be an assessment of 'within-sample' variability.

Turning to the denominator, $\Pr [O_C|H_D, F_C]$, the expert is now required to assign a probability for obtaining the observations made on the perpetrator's images (O_C), given that they are of the left hand of some other, unknown, light-skinned, adult male (H_D) and given the 'conditions' of the perpetrator's images (F_C). While experts can utilise their experience to inform this probability, it would seem that the probability would be more robustly informed by interrogation of data from a relevant database. A database of relevant images could

provide an estimate of the frequency of occurrence of features and thereby inform the probability of observing them in the images of the perpetrator, if the images are not those of the accused.

Step 3 Evaluation of the observations

The expert compared the two sets of images and observed correspondence in the features described and found no differences. How can the probabilities in the likelihood ratio then be assigned?

Looking firstly at the numerator of this equation, $\Pr [O_C|O_D, H_P, F_C, F_D]$, it would be tempting to assign a value of 1 to this probability, as it would seem certain that the perpetrator image would look exactly like it did, given that it was an image of the accused and given the quality of the images. However, the time interval between the two sets of images was unknown. If that period of time had been lengthy, then it may be that the accused's hand could have acquired more scars than are shown in the perpetrator images. If the period had been short, then there would be less opportunity for acquisition of more scars. This uncertainty can be reflected in the assignment of a value less than 1 for the numerator probability. For the purposes of this illustration, however, it is assumed that the numerator probability would be a value close to 1—it seems highly likely that the images of the perpetrator would appear as they do, if they are truly those of the accused and given what is known, or predicted, about 'within-sample' variability in the circumstances of this case.

By comparison, the denominator probability ($\Pr [O_C|H_D, F_C]$) requires expert knowledge of 'between-sample' variability. A database of hand images has previously been reported by one of the authors [14] and consists of images of pairs of hands from 260 individuals of both sexes drawn from a group of UK police officers attending Disaster Victim Identification training at the University of Dundee. The images were taken under good lighting conditions and at high resolution. The location of anatomical features in the images was coded according to a grid system of 24 cells mapped onto each hand image. One hundred and seventy-seven individuals from this database corresponded with the skin colour, gender and age-group features of the perpetrator. Therefore, this subset of the database would appear relevant to inform the denominator probability in this case study. Only two individuals in the subset had two scars in the location (grid cell) of those seen in the case images, and in the opinion of the expert, both of these scars would have been seen if the database images had been taken under the same less than optimal conditions seen in the perpetrator's images (F_C). The two individuals in the database also had no scars visible elsewhere on their left hands. The frequency of occurrence in the subset of the

two scars (and none elsewhere) is therefore 2/177 (1.1 %). Whether this is a robust assessment of the relative frequency in the relevant, larger population is a question that requires some statistical treatment and is one that goes beyond the scope of this paper. We simply wish to illustrate here how the data may be used.

If the relative frequency of 2/177 is an acceptable estimate of the population quantity, then it would seem reasonable to assign a probability of 1.1 %, or 0.011, for obtaining the two scars in that location and none elsewhere.

Substituting the values for the numerator and the denominator provides a likelihood ratio of:

$$LR = \frac{\rightarrow 1}{0.011}$$

or a value of approximately 90.

Following Bayes' theorem (2–5), this value for the LR would increase the prior odds of the images being those of the accused by a factor of approximately 90. To set this number into context, we provide two illustrations.

1. Assume that the prior odds of the images being those of the accused, as assessed by the triers of fact on the basis of the non-scientific evidence, were deemed to be 10:1 against (1/10). The scientific evidence, with a value of 90, would then result in those odds being changed to posterior odds of 9:1 on. Some people prefer to work in terms of probabilities rather than odds, and so, the equivalent values in probability would be 9 % for the prior probability of the images being those of the accused and 90 % for the posterior probability of the images being those of the accused.
2. Assuming, by comparison, the prior odds were 1,000:1 against (1/1,000), the scientific evidence would change that value to posterior odds of 11:1 against (1/11). Converting again to probability, the prior probability of 0.1 % would be changed to a posterior probability of approximately 8 %.

Step 4 Reporting

The AFSP recommended that the likelihood ratio be expressed as a degree of support for one proposition over an alternative. One version of such a verbal scale is provided in the AFSP standards, and following this scale, a value of 90 for the LR would fall towards the upper end of the 'moderate' support category. An evaluative opinion could then be expressed along the lines of:

The finding of a correspondence between the images of the perpetrator's and the accused's left hands provides moderate support for the proposition that the questioned images are those of the left hand of the accused rather than those of an unknown, light-skinned, adult male.

An alternative, valid way of expressing the opinion would be:

The findings... are approximately 90 times more likely to have been obtained if the images are those of the accused rather than of some other, unknown, light-skinned, adult male.

Further development of the LR

So far, in this case, we have only considered the presence of scars when interrogating the database samples. If we were to take into account both the size and the orientation of the scars, we would find that, for the two individuals whose hand images corresponded in number of scars with that of the perpetrator (i.e. two scars at the base of the thumb and none elsewhere), the scars in the two subset samples were of different orientation and size from those seen in the perpetrator image. It would seem reasonable to assume, therefore, that the probability of a 'match', when also taking into account the size and orientation of scars, would be lower than that assigned solely on the presence of scars. This being so, the value of the likelihood ratio would increase when those aspects were also taken into consideration.

To illustrate the potential increase in LR that may occur, assume that the orientation of scars can be classified following a very simple scheme based on the points of a compass. Assume that the image of the back of the hand is orientated in a north–south direction, with the wrist at the south and the tip of the middle finger at the north. Assume that, for each scar in each grid cell, the orientation of a scar can be assigned to one of four orientation categories:

1. North to south
2. East to west
3. North-east to south-west
4. North-west to south-east

Assuming that the orientation of a scar is completely random and that there is no dependence between the orientation of scars and their occurrence in different locations, then the probability that a scar would have a particular orientation would be $\frac{1}{4}$ or 0.25. If this were to be factored into the denominator for this case study, then the probability of seeing two scars in this location, and with the orientations as shown, would be multiplied by a factor of 0.0625 (0.25×0.25). This would give a new denominator value of 0.0006875 (0.011×0.0625), and a new LR of approximately 1,450, an increase in the value of the LR by greater than 1 order of magnitude over the original value. An LR of 1,450 multiplies the prior odds by a factor of 1,450 and falls within the category of 'strong' support [1].

Outcome of case

Purely for sake of completion, as it has no bearing on the scientist's interpretation of the evidence, the accused changed his plea to 'guilty' and was given a public order protection of unlimited duration.

Case 2. The case of the borrowed computer

Step 1 Relevant case circumstances (F_C, F_D)

A man, who was having a relationship with a woman, offered to help her with computer problems. He lent her his computer to use while he fixed hers. The woman found images, some apparently covert of the man's teenage daughter and some of young children, including sexual assault. The police were alerted, and they arrested the man on suspicion of various offences. On arrest, he maintained a stance of 'no comment'. From information visible in the images, the questioned images appeared to have been taken in the accused's house within a year prior to his arrest. The images included the left hand of an adult, light-skinned adult male sexually abusing a young female child. The quality of the offender images was very poor both in terms of lighting and resolution. The images of the accused were taken within days of arrest and under good lighting conditions with optimal resolution.

Step 2 Case assessment

Based on the case circumstances, the principal issue in the case would seem to be the same as that in case 1—are the perpetrator images those of the accused?

It may seem therefore that the same pair of propositions as in case 1 would still be appropriate:

H_P —The image is that of the left hand of the accused.

H_D —The image is that of the left hand of some other, unknown, light-skinned, adult male.

However, in contrast to case 1, the accused has made 'no comment' to the allegation. In this situation, it should be noted that, as expressed above, the alternative proposition H_D is a construction on behalf of the accused. The AFSP standard [1] has offered, based on earlier work [7], various approaches to situations where the accused makes 'no comment'. In this case example, the alternative proposition is based on a default, 'innocent' position. It may be that, nearer to a trial, the accused proposes a different response to the charge. If that were the case, the expert could offer a review and a re-evaluation of their observations.

There are other differences from case 1, and these may influence probability assignments. These include the known time interval between the offence and the taking of the accused's hand images, and the poorer quality of the images.

Preview examination of the perpetrator's images showed two small linear scars on the left hand, one at the base of the

thumb and one at the base of the index finger at either end of the web of skin between the two digits, and no obvious scars elsewhere on the hand (Fig. 2).

Step 3 Evaluation of the observations

On comparison of the two sets of images, good correspondence was observed between the two—the images of the accused showed two small linear scars, one at the base of the thumb and one at the base of the index finger of the left hand, and no scars elsewhere. Considering the numerator of the LR for this case, the expert's view was that it was almost certain the image of the perpetrator (O_C) would have appeared as it did if the images were of the accused's hand (H_P) and given what is known about the accused's images (O_D) and the relevant circumstances (F_C, F_D). It seems reasonable therefore for the expert to assign a value approaching 1, reflecting the expert's judgement of 'almost certain' for the observations.

Turning to the denominator, what is the probability that the perpetrator's images would have shown the observed features (O_C) if they were of the left hand of some unknown, adult, light-skinned male (H_D)? Again, the database mentioned for case 1 would seem to provide a relevant dataset that would provide a relative frequency of occurrence and, hence, a reliable, robust probability for the observations (O_C). When interrogated, the database revealed only one individual with a scar both at the base of the left thumb and at the base of the left index finger, and no scars elsewhere, in 177 relevant samples (0.6 % of left hands). Both scars would most likely have been seen if the database images had been taken under the poorer conditions similar to that of the perpetrator's images (F_C). Therefore, the relative frequency of occurrence in the database of this combination of scars is 1/177 (0.6 %). If that figure is a

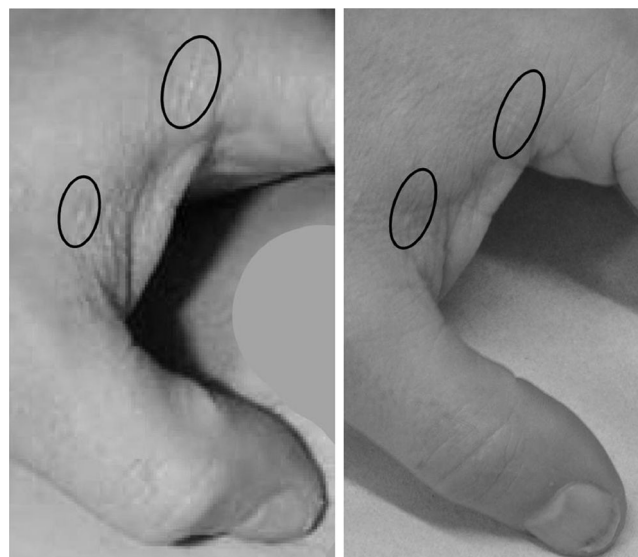


Fig. 2 The perpetrator's image is to the left and that of the accused to the right. Oval outlines show the position of two small scars—one at either end of the digital web between the thumb and the index finger

good assessment of the proportion of hands in the relevant population that would have that combination of scars, then a reasonable assignment for the probability of obtaining the scars in the two locations, and none elsewhere, if the perpetrator's images were not those of the accused's hand (H_D), would be 0.6 % or 0.006.

Substituting the two values, we arrive at a likelihood ratio of:

$$LR = \frac{-1}{0.006} \text{ or a value of approximately } 160.$$

This value for the LR would increase the prior odds of the images being those of the accused by a factor of approximately 160. To illustrate, if the prior odds were deemed to be 10:1 against the images being those of the accused, then the scientific evidence, with a value of 160, would result in posterior odds of 16:1 on. Converting odds to probability, the probability of the images being those of the accused would be changed from 9 to 94 % by the scientific evidence. In comparison, if the prior odds were 1,000:1 against, the scientific evidence would result in posterior odds of 6.25:1 against. Converting again to probability, the prior probability of 0.1 % would be changed to a posterior probability of 16 %.

Because this case example involves scars in two different locations (different grid cells), it raises the issue of whether there is independence in the occurrence of features. However, with such a small dataset and with rare features, it is not possible to assess whether the occurrence of one feature is independent of the occurrence of another. With much larger datasets, it would be possible to test whether features tend to occur together, i.e. show a degree of dependency. With this particular example, we have simply interrogated the dataset for the joint occurrence of the two scars and, in so doing, have automatically taken into account any dependency. Whether this approach provides a robust estimate of the joint occurrence of the two features requires further work.

Step 4 Reporting

As with case 1, following the AFSP document, a value of 160 for the LR could be expressed along the lines of:

The finding of a correspondence between the images of the perpetrator's and the accused's left hands provides moderately strong support for the proposition that the questioned images are those of the left hand of the accused rather than of an unknown, light-skinned, adult male.

or:

The findings... are approximately 160 times more likely to have been obtained if the images are those of the accused rather than of some other, unknown, light-skinned, adult male.

Again, as with case 1, the LR would be increased if the size and orientation of scars were to be included in the evaluation. The two individuals who 'match' in the database

have scars in different orientations, and of different size, from the perpetrator's scars.

Case outcome

The accused changed his plea to guilty and is awaiting sentencing.

Case 3. The case of the supermarket photographer

This case illustrates a treatment for cases in which there is no match within a database or its subset.

Step 1 Relevant case circumstances (F_C, F_D)

Staff of the photographic department of a supermarket noticed pornographic images on a memory card that a male customer had brought in for processing. The images showed an adult, light-skinned male, sexually assaulting an apparently unconscious, adult female. The supermarket alerted police and the customer was arrested. The woman in the pictures alleged that the man had told her that he was a professional photographer and she accompanied him to his home. He plied her with alcohol and drugs. She was unconscious for a time, and when she awoke, she was naked, and the man was standing over her. The time interval between the incident and the arrest of the man was approximately 14 days. The reference images of the suspect's hands were taken under good lighting conditions approximately 4 months after the incident. The lighting conditions of the questioned images were quite poor, being taken in the house of the accused, and the resolution was also quite low. At court, the accused pleaded 'not guilty' to the offence of indecent assault.

Step 2 Case assessment

Preview of the perpetrator's images (O_C) showed the left hand of a light-skinned adult male, and therefore, the approach applied to our first two case studies was also applicable to this case, with the pair of propositions being:

H_P —The image is that of the left hand of the accused.

H_D —The image is that of the left hand of some other, unknown, light-skinned, adult male.

Step 3 Evaluation of the observations

On comparison, correspondence was observed between the two image sets, both showing a large, non-linear scar in the region of the proximal phalanx of the index finger and no scars elsewhere (Fig. 3). Considering the numerator of the LR for this case, the expert may consider it extremely probable that the image of the perpetrator (O_C) would appear as it did given the images are of the accused's left hand (H_P). It seems reasonable therefore to assign a value again approaching 1.

Turning to the denominator, what is the probability that the perpetrator's images would show the observed features (O_C) if

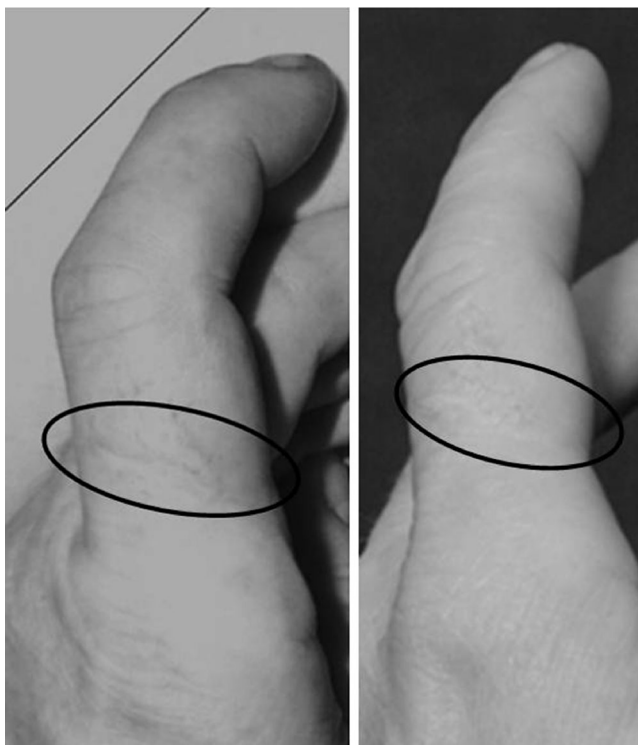


Fig. 3 The perpetrator's image is to the *left* and that of the accused, to the *right*. *Oval outlines* show the position of a large irregular scar across the width of the index finger

they were of the left hand of some unknown, adult, light-skinned male (H_D)? Again, the database mentioned for cases 1 and 2 would seem to provide a relevant subset of data that would provide a relative frequency of occurrence and, hence, a reliable and robust probability for the observations (O_C). However, when interrogated, the database revealed no individuals with a large, non-linear scar on the skin of the proximal phalanx of the index finger. This then poses the problem of how to assign a probability for the feature when the frequency of occurrence of that feature in a database is zero.

This is a problem that has been addressed elsewhere in forensic science, and while there are various ways of estimating a relative frequency of occurrence of features that have not been seen within datasets [15, 16], perhaps the simplest way is to add the case-specific instances into the dataset and use the 'new' dataset to calculate a relative frequency of occurrence. In the DNA-profiling world, this is known as 'size-bias correction' [15]. Following this procedure, in this case example, the scientist could proceed as follows.

As it is the denominator probability that is being considered, it is a condition (it is a given) that the perpetrator and the accused are not the same person. On this assumption, it can be said that two new instances of the feature have been observed, and these two examples should then be added to the dataset. In this case study, adding two entries to the database for this feature in that specific anatomical location results in a relative

frequency of occurrence of 2 in 179 (177 males plus the two new entries) or 1.1 %. If this is accepted as a reasonable estimate, then a value of 0.011 could be assigned to the probability of obtaining one, non-linear scar in this anatomical location if the images were those of the left hand of an unknown, light-skinned, adult male. Substituting the values for the numerator and denominator, we arrive at the same likelihood ratio as for case 1, i.e. an LR of approximately 90.

A more realistic assessment of the relative frequency of occurrence of this and other features could be obtained through adding more samples to the database. The question would be 'how many more?' This is not an easy question to answer but would depend on the rarity of the features in question. If they truly were very rare, then a very large dataset may be required. Alternatively, if they were relatively common, a small database would suffice. Further discussion of this problem is beyond the scope of this paper.

As with the previous two case studies, there is the potential to enhance the value of the LR if size, shape and orientation of the scar could be taken into account. However, there is no occurrence within the dataset of a scar in this location. It is therefore not possible, based solely on this database, to assess the frequency of occurrence of the size, shape and orientation of scars in this location. It would be possible to rely on an expert's opinion on this frequency, but the validity of that opinion would depend on the depth of the expert's experience and on their recall of such scars. For the purposes of this case study, we have not attempted to extend the evaluation of the LR for such additional characteristics. The importance of the availability of large datasets to help estimate frequencies of occurrence of features cannot be overstated.

Step 4 Reporting

The reporting of the LR for this case would be along similar lines to that for case 1:

The finding of a correspondence between the images of the perpetrator's and the suspect's left hands provides moderate support for the proposition that the questioned images are those of the left hand of the suspect rather than those of an unknown, light-skinned, adult male.

or:

The findings... are approximately 90 times more likely to have been obtained if the images are those of the accused rather than of some other, unknown, light-skinned, adult male.

Outcome of case

Initially, the accused pleaded guilty to the offence of indecent assault, but following the influences of the judgement in *Cadder v HM Advocate* [17], he changed his plea to 'not

guilty'. He was subsequently found guilty by jury and sentenced to 4 years in prison with a 3-year extended licence.

Summary and conclusions

This paper has demonstrated an application of the principles of logical and robust interpretation of expert evidence in cases involving the comparison of scars visible in images of hands. It has shown how likelihood ratios may be evaluated through the use of databases of images to help assign subjective, expert probabilities for the occurrence of scars. Examples have been presented of the way in which expert opinions, based on an evaluation of likelihood ratios, may be expressed.

This paper has indicated the value of the interpretation of the presence and location of scars as a uni-modal anatomical feature of identity and whilst it has been shown to be of practical value, the discriminatory capacity of this approach to image analysis will likely be significantly enhanced when additional features are incorporated into a multi-modal response.

In addition, further work is required to increase the size of databases so that:

1. The occurrence of rare features may be assessed more realistically
2. The degree of dependency of features can be more clearly understood
3. The database covers other populations that may be relevant to forensic issues

Finally, following the examples provided by other areas of forensic expertise [18–20], we suggest that experiments should be performed to calibrate the performance of experts in the comparison of hand images.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

1. Association of Forensic Science Providers (2009) Standards for the formulation of evaluative forensic science expert opinion. *Sci Justice* 49:161–164
2. Lindley DV (1991) Probability. In: Aitken CGG, Stoney DA (eds) *The use of statistics in forensic science*. Ellis Horwood, Chichester, pp 27–50
3. Robertson B, Vignaux GA (1995) *Interpreting evidence*. Wiley, Chichester
4. Aitken CGG, Taroni F (2004) *Statistics and the evaluation of evidence for forensic scientists*. Wiley, Chichester
5. Aitken CGG, Roberts P, Jackson G (2010) *Fundamentals of probability and statistical evidence in criminal proceedings*. Royal Statistical Society, London
6. Jackson G (2009) Understanding forensic science opinions. In: Fraser J, Williams R (eds) *Handbook of forensic science*. Willan, Cullompton, pp 419–445
7. Jackson G, Jones S, Booth G, Champod C, Evett IW (2006) The nature of forensic science opinion—a possible framework to guide thinking and practice in investigations and in court proceedings. *Sci Justice* 46:33–44
8. Champod C, Evett IW, Jackson G (2004) Establishing the most appropriate databases for addressing source level propositions. *Sci Justice* 44:153–264
9. Evett IW, Jackson G, Lambert JA, Mc Crossan S (2000) The impact of the principles of evidence interpretation on the structure and content of statements. *Sci Justice* 40:233–239
10. Taroni F, Aitken CGG, Garbolino P (2001) DeFinetti's subjectivism, the assessment of probabilities and the evaluation of evidence: a commentary. *Sci Justice* 41:145–150
11. Lucy D (2005) *Introduction to statistics for forensic scientists*. Wiley, Chichester, pp 4–6
12. Cook R, Evett IW, Jackson G, Jones PJ, Lambert JA (1998) A model for case assessment and interpretation. *Sci Justice* 38:151–156
13. Jackson G, Jones PJ (2009) Case assessment and interpretation. In: Jamieson A, Moenssens A (eds) *Wiley encyclopedia of forensic science*. Wiley, Chichester, pp 483–497
14. Black S, MacDonald-McMillan B, Mallett X (2013) The incidence of scarring on the dorsum of the hand. *Int J Leg Med*. doi:10.1007/s00414-013-0834-7
15. Curran JM, Buckleton JS, Triggs CM, Weir BS (2002) Assessing uncertainty in DNA evidence caused by sampling effects. *Sci Justice* 42:29–37
16. Aitken CGG (1991) Populations and samples. In: Aitken CGG, Stoney DA (eds) *The use of statistics in forensic science*. Ellis Horwood, Chichester, pp 51–82
17. Cadder v HM Advocate (2010) http://www.supremecourt.gov.uk/docs/UKSC_2010_0022_Judgment.pdf. Accessed 26 October 2012
18. Neumann C, Evett IW, Skerrett J (2012) Quantifying the weight of evidence from a forensic fingerprint comparison: a new paradigm. *J R Stat Soc, Ser A* 175:371–415
19. Junod S, Pasquier J, Champod C (2012) The development of an automatic recognition system for earmark and earprint comparisons. *Forensic Sci Int* 222:170–178
20. Morrison GS (2011) Measuring the validity and reliability of forensic likelihood-ratio systems. *Sci Justice* 51:91–98