WEIGHING EVIDENCE: QUANTITATIVE MEASURES OF THE IMPORTANCE OF BITEMARK EVIDENCE

J.M. Kittelson,[1] J.A. Kieser,[2] D.M. Buckingham,[3] and G.P. Herbison[4]

1. *Department of Preventive Medicine and Biometrics, University of Colorado Health Sciences Center, Denver, Colorado, USA*
2. *Department of Oral Sciences and Orthodontics, University of Otago, Dunedin, New Zealand*
3. *Department of Law, University of Otago, Dunedin, New Zealand*
4. *Department of Preventive and Social Medicine, Dunedin School of Medicine, University of Otago*

ABSTRACT

Quantitative measures of the importance of evidence such as the "likelihood ratio" have become increasingly popular in the courtroom. These measures have been used by expert witnesses formally to describe their certainty about a piece of evidence. These measures are commonly interpreted as the amount by which the evidence should revise the opinion of guilt, and thereby summarize the importance of a particular piece of evidence. Unlike DNA evidence, quantitative measures have not been widely used by forensic dentists to describe their certainty when testifying about bitemark evidence. There is, however, no inherent reason why they should not be used to evaluate bitemarks. The purpose of this paper is to describe the likelihood ratio as it might be applied to bitemark evidence. We use a simple bitemark example to define the likelihood ratio, its application, and interpretation. In particular we describe how the jury interprets the likelihood ratio from a Bayesian perspective when evaluating the impact of the evidence on the odds that the accused is guilty. We describe how the dentist would calculate the likelihood ratio based on frequentist interpretations. We also illustrate some of the limitations of the likelihood ratio, and show how those limitations apply to bitemark evidence. We conclude that the quality of bitemark evidence cannot be adequately summarized by the likelihood ratio, and argue that its application in this setting may be more misleading than helpful. **(J Forensic Odontostomatol 2002;20:31-7)**

Keywords: bitemark, Bayesian, likelihood ratio

INTRODUCTION

The thrust of modern bitemark analysis has generally been to resolve mechanistic questions about reproduction, recording and comparison procedures to be followed.[1] Of secondary importance have been questions about the uniqueness of the dentition and about how bitemark evidence is to be presented in the courtroom. Progress in resolving questions about the individuality of tooth shape, size and position[2-6] has been cautious, perhaps because investigators have been loathe to consider dental development from a population perspective. As a consequence, a comprehensive characterization of dental individuality comparable to that of DNA data has not emerged.

The presumption of dental uniqueness is the *raison d'etre* for the admissibility of bitemark evidence.[7,8] Yet if one accepts that no two bites by the same biter will be identical in all ways, and that the principle of individuality implies that no two bitemarks are identical, the question becomes with what confidence can we distinguish between two people's bites on the same surface, and can this confidence be meaningfully quantified?

Unlike DNA evidence, bitemark analysis does not have a quantitative base; thus, it is crucial for the forensic dentist to understand how to relate an inherently qualitative assessment to the quantitative measures used in other fields. Because it is the duty of the expert witness to inform the jury, the goal is to give an appropriate interpretation of certainty of that expert opinion rather than develop a technological black box that produces a number which substitutes for that appropriate interpretation.

Quantitative measures of the importance of a piece of evidence are based on the idea that the perception of guilt can be described using probability. In a criminal trial the jury is instructed to evaluate the evidence, and in that process must assume innocence unless guilt is established "beyond a reasonable doubt." This basic judicial premise implies that the degree of certainty must be a central consideration, and that a guilty decision should only be reached when the probability of innocence is sufficiently small. Jury deliberations can be viewed as the process of determining the probability of guilt, and a "reasonable doubt" could be phrased in terms of this probability. Although it is debatable whether or not formal quantification of uncertainty is helpful to juries, its use during trials is increasing - especially its application to evidence offered by expert witnesses. As a consequence it is important for expert witnesses to develop an understanding and intuition for the measures of uncertainty as applied in the courtroom.

The primary objective of this paper is to describe the application, calculation and interpretation of quantitative measures of the importance and reliability of evidence in forensic dentistry. Since the seminal work of Lindley[10] and Evett[11] there has been a growing interest in the "Bayesian approach" to the quantification and evaluation of expert forensic testimony.[9,12,13] Recent papers by Malokoff,[14] Goodman[15] and others are further evidence for a renewed interest in the Bayesian approach. The principles behind these measures are somewhat abstract, and as a result there is an impression that forensic dental evidence may have become overshadowed by sophisticated and mathematically complex techniques which are not easily understood.[16,17] In fact these measures are all directed at expressing the probability of guilt, and in particular, how that probability changes with new evidence. In what follows, we review these measures and the fundamental ideas from which they derive in order to describe their appropriate interpretation and application to the evaluation of human bitemark evidence.

Quantitative Evaluation of Evidence

A new piece of evidence in a trial revises the certainty or probability of guilt. The amount by which the probability is revised depends on the importance and the quality of the evidence. The jury relies on the expert witness to provide this information so that the evidence has an appropriate impact on the verdict. In this section we describe a formal framework for the interpretation and evaluation of evidence in a trial using a numeric rating know as the "likelihood ratio." We first describe the likelihood ratio as a quantity used to interpret the evidence from the perspective of the jury and then describe the same quantity as a measure used by the expert witness to evaluate the reliability of the evidence. We note that likelihood ratios are sometimes considered synonymous with "Bayesian reasoning" in the courtroom. In the following discussion we describe what this means, and contrast this interpretation with the "frequentist" interpretation used by the expert witness to derive the likelihood ratio.

The jury and the Bayesian interpretation

Conceptually, at the beginning of the trial (in the absence of any evidence and under the presumption of innocence) the probability of guilt should be very small. Initially the probability should be approximately $1/N$ where N is the largest possible number of potential perpetrators (e.g., N is the number of people in the city where the crime occurred). This could be interpreted as the probability that the accused individual has ended up on trial simply because the police randomly selected one individual from this population. It is equivalent to expressing this probability using the "odds." The use of odds originates in wagering where it is the amount of money required for a fair bet. Odds can be calculated from the probability: if the probability of guilt is p, then the odds of guilt are $p/(1-p)$. In a finite sample where p is the proportion of people with a particular trait, then the odds of the trait is the ratio between the number of people with the trait and the number of people without the trait. In our courtroom application, if the initial probability of guilt is $1/N$, then the odds of guilt are:

$$\frac{\frac{1}{N}}{1 - \frac{1}{N}} = \frac{1}{N - 1} \qquad \text{eq 1}$$

The evidence in the trial is used to either increase or decrease the initial probability (or odds) so that a guilty or not-guilty verdict can be returned. One way to interpret the impact of a piece of evidence on the trial, and hence its importance, is to measure how much it revises the probability of guilt. The likelihood ratio (LR) is the commonly used measure of the impact of a piece of evidence on the perception of guilt. It is defined as the ratio between the odds of guilt after the evidence is considered to the odds of guilt prior to its introduction:

$$LR = \frac{\textit{odds of guilt given the evidence}}{\textit{odds of guilt prior to evidence}}$$

If the likelihood ratio for a piece of evidence is less than 1.0, then that evidence has decreased the odds of guilt. If it is greater than 1.0, then the odds of guilt are larger than they were without the evidence.

To illustrate, we consider a hypothetical example involving the trial of a man for an assault at a party. Suppose that during the trial it has been clearly established that the assailant was male and attended the party along with 49 other males. Suppose that there are no witnesses, so the guilt or innocence of the defendant will be decided on circumstantial evidence. Furthermore, suppose that the victim was bitten by the assailant, and that a forensic dentist has been asked to evaluate the bitemark and bite casts for all male party goers. On the witness stand, the dentist states that the most prominent feature of the bitemark is the mesiodistal incisor width which is at least 9.5mm, and that the bite casts show that 3 of the 50 male party goers (including the accused) have incisor width exceeding 9.5mm. In this example the prior odds that the accused is guilty are 1/49 (which corresponds to a probability of 1/50), and after the bitemark evidence those odds increase to 1/2 (corresponding to a probability of 1/3). The likelihood ratio calculated using the pre- and post-evidence odds is $1/2 \div 1/49 = 24.5$. From the jury's perspective, the bitemark evidence has resulted in a 24.5-fold increase in the odds of guilt.

A central requirement for the above calculation is the presence of a clearly defined group of 50 suspects - something which may not be available. Consider a second example in which an assault occurred on the streets of a large city, and the only clearly established prior information is that the assailant is male. In this example we can denote the size of the male population by N, but it cannot be directly calculated. Once again, suppose that the dentist determines that the assailant has a mesio-distal incisor width of at least 9.5mm, and that the accused satisfies this condition. Suppose further that population studies indicate that mesiodistal incisor width among males follows a Normal distribution with a mean of 8.7mm and a standard deviation of 0.4mm; that is, the assailant's incisors are at least 2 standard deviations above the mean. From the properties of the Normal distribution, only 2.3% of the population is more than 2 standard deviations above the mean. It follows that the incisor width evidence has reduced the size of the population of potential perpetrators from N to $0.023 \times N$. In this example the prior odds of guilt are $1/(N - 1)$ and the revised odds are $1/(0.023\,N - 1)$, so the likelihood ratio (after simplification as in equation 1) is:

$$LR = \frac{1/(rN - 1)}{1/(N - 1} = \frac{1 - \frac{1}{N}}{r - \frac{1}{N}}$$

where r is the prevalence of an evidential characteristic in the N possible perpetrators. In the above example, $r = 0.023$ and if N is large, $LR \approx 1/0.023 = 43.5$. The incisor evidence has produced a 43.5-fold increase in the odds of guilt.

Note that the likelihood ratio requires knowledge of r and N. In the first example both could be determined by actual measurements taken on all possible perpetrators, which would clearly be impossible in most situations. In the second example it is sufficient to know that N is large as long as there is secondary information about the distribution of the bitemark in the relevant population. The second approach is commonly used when there is a natural estimate of the prevalence of an evidential trait. A good example is with DNA testing where the probability of a match in a randomly selected individual can be determined from the number of independent alleles examined. As discussed below, an incorrect likelihood ratio could be obtained if the large population approach were applied to a small group of potential perpetrators or *vice versa*.

The likelihood ratio as defined and interpreted above is a Bayesian quantity because it measures the probability of truth (i.e., guilt) given the data (i.e., evidence). A Bayesian interpretation contrasts with a frequentist interpretation which is based on measuring the probability of the data (evidence) given truth (innocence or guilt). The relative merits of Bayesian and frequentist approaches to statistical analysis is currently the subject of debate. Advocates of frequentist approaches prefer methods that are not affected by an arbitrary prior distribution (a technical necessity for the Bayesian approach), whereas advocates of Bayesian methods prefer the interpretability of a Bayesian analysis. In fact the Bayesian/frequentist argument does not really apply to the interpretation of courtroom evidence. The likelihood ratio has a meaningful frequentist interpretation in addition to its Bayesian interpretation. We now show that a frequentist interpretation is more useful to the expert witness than the Bayesian interpretation used by the jury.

The expert witness and the frequentist evaluation

The expert witness must provide the jury with some indication of the certainty or uncertainty in his/her conclusions about the evidence. When the determination is subjective (as is the case with much of forensic dentistry), then the certainty must be conveyed verbally. When quantitative determinations are possible, then the likelihood ratio provides a more precise measure of certainty. To calculate a likelihood ratio by quantifying the odds of guilt as described above would be inappropriate for an expert witness. The jury must assess guilt; the expert witness must evaluate a particular piece of evidence.

The likelihood ratio can be re-expressed using elementary probability relationships so that the expert witness can report the same likelihood ratio without quantifying the pre- and post-evidence odds of guilt:

$$LR = \frac{\textit{odds of guilt given the evidence}}{\textit{odds of guilt prior to evidence}}$$

$$= \frac{\textit{probability of evidence given guilty}}{\textit{probability of evidence given not guilty}}$$

The first interpretation of the likelihood ratio is the Bayesian measure described above. The second interpretation is a frequentist measure because it describes the probability of the evidence under different assumptions about guilt .

Once again, consider the first example of the assault at the party. Suppose that the dentist has finished testimony and the lawyer for the defence asks for the likelihood ratio as a measure of the importance of the evidence. During evaluation of the bite casts and bitemark the dentist needed to ask two questions: (1) how likely are various bite characteristics in the assailant (the person who made the bitemark - not necessarily the accused), and (2) how likely are those characteristics in the non-guilty population? The dentist must assess the bitemark for one or more distinctive characteristics; i.e., those characteristics which are both readily apparent in the bitemark and uncommon in the non-guilty population. Assigning probabilities to these assessments leads to the alternative form for the likelihood ratio. In the party assault example the dentist would have examined the bitemark and decided that the assailant's incisor width could be measured with certainty, and that wide incisors are relatively rare in the non-guilty population. Expressed as probability, the dentist would be certain that the assailant had an incisor width of at least 9.5mm (i.e., the probability of width $\geq$ 9.5mm in the guilty person is 1.0), and from the casts would measure that 3 of the 50 potential suspects had wide incisors (which implies that 2 of 49 non-guilty individuals had wide incisors). According to the alternative version of the likelihood ratio:

$$LR = \frac{\textit{probability of evidence given guilty}}{\textit{probability of evidence given not guilty}}$$

$$= \frac{1/1}{2/49} = 24.5$$

which is the same as the value derived above by quantifying the odds of guilt.

If, as in the second example, the assault occurred in a large city, it would only be known that the number of potential assailants is very large. Once again, the dentist would select a characteristic which was both readily apparent in the bitemark and rare in the non-guilty population. As in the party example, the incisor width would be a good characteristic if width is readily measured in the bitemark and if the measurement was so extreme as to make it unusual in the general population. If, as above, there are N individuals in the population of potential assailants (1 guilty and N-1 not guilty), and the proportion with incisors larger than 9.5mm is r, then the probability of the wide incisors in the non-guilty portion of the population is $(rN-1)/(N-1)$, and the likelihood ratio is:

$$LR = \frac{1}{(rN - 1)/(N - 1)} = \frac{1 - \frac{1}{N}}{r - \frac{1}{N}}$$

Once again, the likelihood ratio is approximately $1/r$ when N is large, and if $r = 0.023$, then $LR = 43.5$ as obtained above when quantifying the odds of guilt.

The frequentist interpretation of the likelihood ratio reflects the difference in definition. In the party assault, the dentist would report that the probability that the assailant has wide incisors is 24.5 times the probability of large incisors in the non-guilty population. The jury would interpret this as the amount by which the pre-evidence odds of guilt have increased. Similarly, in the city assault the dentist reports that the probability of wide incisors in the guilty individuals is 43.5 times the probability of wide incisors in a non-guilty individual; the jury interprets this as change in the pre-evidence odds of guilt.

In the previous section we motivated the likelihood ratio as the jury's *interpretation* of how the evidence changes their perception of guilt. The forensic dentist focuses on the *evaluation* of the evidence, and therefore focuses on its quality. In this role the dentist is interested in unusual characteristics that identify a distinct bitemark which serves to narrow the list of possible suspects. By definition unusual characteristics are rare, and rare characteristics will make high quality bitemark evidence. Thus, the probability of a bite characteristic in the general population is a natural measure for evaluation of the certainty of the evidence. In this regard we suggest that the jury views the likelihood ratio from the Bayesian perspective, and the expert witness views it as a frequentist evaluation.

One issue which is readily apparent in the frequentist interpretation, which is less apparent in the Bayesian interpretation is the role of certainty in characterising a bitemark trait. We assumed in the incisor width example that the dentist was certain that the assailant's incisor width was at least 9.5mm. This assumption was explicit in the frequentist interpretation where the numerator was the probability that the incisor width in the assailant exceeded 9.5mm. If the bitemark evidence was poor, then the dentist may not be certain of this characteristic, and the probability could be chosen to be less than 1.0. In contrast, uncertainty in the characterisation of the bitemark is not as explicit in the Bayesian interpretation where it would alter the odds of guilt given the bitemark characteristic. Once again, the frequentist interpretation is more natural for the dentist who must incorporate uncertainty in the bitemark trait into their assessment.

Limitations of Quantitative Measures

The simple examples given above treat the likelihood ratio as if it is a fixed property that is directly determined by the quality of the evidence. In fact, the likelihood ratio depends both on the size of the population of possible perpetrators (N) and on the prevalence of the evidential trait (r) in that population. It follows that the likelihood ratio will change according to what information is already known at the time the evidence is presented. For example, if N decreases during the course of the trial (as we might expect) then the likelihood ratio will tend to increase as long as r stays constant. For many types of evidence the prevalence of the evidential characteristic, r, is different in different reference populations; thus for example, if we know that the assailant had large feet, then it would not be so unlikely that they would also have large teeth in which case r will be larger and the LR smaller.

We return to the second example to illustrate the interplay between different pieces of evidence and their effect on the likelihood ratio. Suppose that in the large city assault, the gender of the assailant was not clear, so that the initial population included both females and males. Females have smaller teeth, and suppose the probability that incisor width exceeds 9.5mm in a population of females is 0.0062. If the population is equally split between males and females, then the overall probability that the incisor width exceeds 9.5mm is the average of the male and female probabilities, (0.023 + 0.0062)/2 = 0.0146. Following the same calculations given earlier, the likelihood ratio when the population of possible assailants includes both males and females is approximately 1/0.0146 = 68.5, as compared to 1/0.023 = 43.5 if the assailant is known to be male. Thus, the likelihood ratio may change as the population of possible perpetrators is refined. This can happen both at the initiation of the trial, and if the population is revised by other evidence offered as part of the trial. The likelihood ratio will not change if the particular piece of evidence is completely independent of any other evidence; for example, if the prevalence of large teeth is the same in both males and females, then the likelihood ratio for the incisor width evidence would be the same regardless of whether it followed or preceded any gender evidence.

In many trials the list of potential perpetrators is actually quite small, and as in the party assault example, the evidence is used to select the actual perpetrator from a short list. In this setting it would be inappropriate to use the large population methods described in the second example. For example, suppose that a DNA analyst reports a one in 1,000,000 chance of a match with a randomly selected individual from a large population. If there are only 10 possible perpetrators, then the likelihood ratio cannot be approximately 1,000,000 as it would be in the larger population. If all 10 suspects have DNA profiles and the defendant is the only match, then the likelihood ratio is infinite (= 1/0 ÷ 1/9). If however the DNA matches 3 of the 10 suspects (something which might happen if suspects are related) then the likelihood ratio is 3.0 (= 1/3 ÷ 1/9). In fact bitemark evidence may be most useful in these small-population situations where it is possible to match the mark to a small number of bite casts.

One of the theoretically appealing aspects of the likelihood ratio is that the overall likelihood ratio for large blocks of evidence (or for all evidence) can be calculated by taking the product of all likelihood ratios from the individual pieces of evidence. Unfortunately, this theoretical possibility is impossible in practice. Practical problems arise because some evidence is not amenable to quantification; thus, even with quantifiable evidence, the likelihood ratio will depend on earlier evidence for which quantification may be impossible. Similarly, even if likelihood ratios can be calculated for all previous evidence, it may not be possible to define explicitly the size of the population of potential perpetrators, which as described above can have important effects on the likelihood ratio. Finally, even if a reasonable likelihood ratio can be calculated for every piece of evidence in a trial, it is questionable as to whether or not the quantitative measure would lead to better decisions. One attempt to calculate an overall likelihood ratio in a trial resulted in a reversal of a verdict on the basis that the jury had been misdirected.[18]

DISCUSSION

Interpretation of bitemark evidence is not a quantitative process. Although some characteristics can be directly measured (e.g., tooth dimensions, intercanine distances), the ability to discern measurable characteristics depends very much on the circumstances surrounding the bite; bitemarks are not usually made to facilitate later evaluation. Furthermore, unlike DNA testing, there is no theoretical basis for determining the population prevalence of bitemark characteristics, so even if measurements can be made their distribution (prevalence) in the reference population is probably unknown.

We conclude that bitemark evidence is inherently qualitative, and the use of quantitative measures to describe the importance of bitemark evidence would be misleading. Such quantitative measures must derive from quantitative justification, and basing a likelihood ratio on professional opinion or experience is likely to give a misleading impression of the importance of the evidence and a false sense of objectivity to a subjective determination. This point has recently been emphasised by Taroni et al.,[19] who noted that frequentist probabilities were objective and Bayesian probabilities were subjective. That is, frequentists rely upon long-run repetitions of an observational event under defined and constant conditions. In contrast, Bayesian or subjective probabilities refer to the level of belief that an expert may hold, based on his or her experience, knowledge and information, about a single event whose falsity or truth is unknown.

Although further studies might determine the population prevalence for common bitemark characteristics, calculation of likelihood ratios for bitemark evidence will always be complicated by the variable nature of the circumstances surrounding the bite. This is not to say that bitemark evidence is useless in a courtroom. In fact bitemark evidence may be excellent for selecting the perpetrator from a small group of suspects (as illustrated in the party assault example). However even in these situations the jury may be better informed by a careful explanation of the bite characteristics than a formal calculation of a likelihood ratio.

It is essential that expert witnesses have some understanding of quantitative measures of the importance of evidence. Their use is common, and their apparent objectivity is appealing to judges and lawyers. In some circumstances measures such as the likelihood ratio greatly facilitate the interpretation of a piece of evidence, however inappropriate use serves only to confuse matters. When it comes to bitemark evidence, we recommend that the forensic dentist understand the likelihood ratio, and be able to offer an explanation (based on issues described above) of why it may not apply to the case at hand.

REFERENCES

1. Naru AS. Method for the analysis of human bite marks. Forensic Sci Rev 1997;20:124-34.
2. Taylor RMS. Variation in form of human teeth: I An anthropological and forensic study of maxillary incisors. J Dent Res 1969;48:173-82.
3. Taylor RMS. Variation in form of human teeth: II An anthropological study of maxillary canines. J Dent Res 1969;48:173-82.
4. Kieser JA. Human adult odontometrics. Cambridge: Cambridge University Press, 1990.
5. Sognnaes RF, Rawson RD, Gratt BM, Nguyen NBT. Computer comparison of bitemark patterns in identical twins. J Am Dent Assoc 1982;105:449-51.
6. Rawson RD, Ommen RK, Kinard G, Johnson J, Yfantis A. Statistical evidence for the individuality of human dentition. J Forensic Sci 1984;29:245-53.
7. McKenna CJ, Haron MI, Taylor JA. Evaluation of a bitemark using clear acrylic replicas of the suspect s dentition. J Forensic Odontostomatol 1999;17:40-3.
8. Court of Criminal Appeal (Qld); Raymond John Carroll. 1985; 19A Crim R.
9. Robertson B, Vignaux GA. Interpreting evidence. Chichester: John Wiley, 1996.
10. Lindley DV. A problem in forensic science. Biometrika 1997;64:207-13.
11. Evett IW. Bayesian inference and forensic science: problems and perspectives. The Statistician 1987;36:99-105.
12. Good IJ. Weight of evidence and the Bayesian likelihood ratio. In: Aitken CGG, Stoney DA Eds. The use of statistics in forensic science. New York: Ellis, 1991.
13. Foster KR, Huber PW. Judging science: scientific knowledge and the Federal Courts. Cambridge: MIT Press, 1999.
14. Malakoff D. Bayes offers a new way to make sense of numbers. Science 1999;286:1460-4.
15. Goodman SA. Toward evidence-based medical statistics: the p-value fallacy. Anns Int Med 1999;130:995-1013.
16. Clement JG. ASFO guest editorial. Forensic Odontology News 1999;18:3-12.
17. Taroni F, Mangin P, Perrior M. Identification concept and the use of probabilities for forensic odontology. J Forensic Odontostomatol 2000;18:15-7.
18. R v. Adams; Court of Appeal of England and Wales (Criminal Division), Royal Courts of Justice, The Strand London WC2.
19. Taroni F, Aitkin CGG, Garbolino. De Finetti s subjectivism, the assessment of probabilities and the evaluation of evidence: a commentary for forensic scientists. Science and Justice 2001;145-50.

Address for correspondence:
Professor Jules Kieser
Department of Oral Sciences and Orthodontics
School of Dentistry
PO Box 647
Dunedin, New Zealand
E-mail: jules.kieser@stonebow.otago.ac.nz