Available online at www.sciencedirect.com
ScienceDirect
Journal of Forensic and Legal Medicine xxx (2008) xxx–xxx
JOURNAL OF FORENSIC AND LEGAL MEDICINE
www.elsevier.com/jflm

Review

Forensic Epidemiology: A systematic approach to probabilistic determinations in disputed matters

Michael D. Freeman PhD, MPH, DC (Adjunct Associate Professor of Forensic Medicine and Epidemiology, Clinical Associate Professor) [a,b,*], Annette M. Rossignol ScD (Professor) [c], Michael L. Hand PhD (Professor) [d]

[a] *Institute of Forensic Medicine, Faculty of Health Sciences, University of Aarhus, 205 Liberty Street, Suite B, Salem, OR 97301, USA*
[b] *Department of Public Health and Preventive Medicine, Oregon Health and Science University School of Medicine, USA*
[c] *Department of Public Health, Oregon State University, USA*
[d] *Atkinson Graduate School of Management, Willamette University, USA*

Received 18 March 2007; received in revised form 15 October 2007; accepted 13 December 2007

Abstract

Forensic medicine testimony often relies upon terms of probability to enhance the strength of the testimony. Such terms must have a demonstrably reliable and accurate basis; otherwise their use is speculative, unjustified, and potentially harmful. Forensic Epidemiology is introduced as a framework from which probabilistic testimony can be assessed in settings in which it is either proffered or encountered. In this paper, common forensic uses of probability are reviewed, appropriate methods for presenting such testimony are proposed, and inappropriate uses of probability and epidemiologic concepts and data, as well as a logical fallacies commonly observed in forensic settings are presented. A previously unpublished logical fallacy, the "Prior Odds" Fallacy, is also introduced.

Keywords: Probability; Epidemiology; Forensic; Prior Odds Fallacy; Sensitivity; Specificity; Positive Predictive Value

1. Introduction

In 1999, British solicitor Sally Clark was convicted in a United Kingdom court of murdering two of her children. Both of the infants died within weeks of birth under circumstances originally diagnosed by some experts as sudden infant death syndrome (SIDS) or cot death.[1] An important witness for the prosecution was the prominent pediatrician Sir Roy Meadow, who testified that the probability of two cot deaths in one family was exceedingly remote; about 1 in 73,000,000. The miniscule probability that the deaths were due to natural causes was used by the prosecution as evidence that the deaths were homicidal. Meadow was a well-known and often-used prosecution witness in similar proceedings, having been the first to promulgate the concept of Munchausen Syndrome by Proxy (MSbP) in which a parent injures or sickens a child as a means of procuring medical attention.[2] Meadow's Law, a heuristic attributed to Meadow that pertains to multiple cot deaths in families, states that unless otherwise proven, one death is tragic, two is suspicious, and three is murder.[3]

In the Clark case, the estimate of 1 in 73,000,000 was derived from squaring the observed risk of a single cot death in an affluent non-smoking family; estimated at 1 in 8500. Meadow's testimony created a furor among statisticians, with the president of the Royal Statistical Society

* Corresponding author. Address: Institute of Forensic Medicine, Faculty of Health Sciences, University of Aarhus, 205 Liberty Street, Suite B, Salem, OR 97301, USA. Tel.: +1 503 586 0127; fax: +1 503 586 0192.
E-mail address: forensictrauma@gmail.com (M.D. Freeman).

1752-928X/$ - see front matter © 2008 Elsevier Ltd and FFLM. All rights reserved.
doi:10.1016/j.jflm.2007.12.009

writing an open letter of complaint to the Lord Chancellor,[4] and the British Medical Journal publishing an editorial concerning the error of allowing non-qualified experts to testify regarding unsound statistically based arguments.[5] The primary complaint with Meadow's testimony was that from a statistical perspective, he treated the specific risk to the Clark family for cot death like they were a randomly selected family with no predilection for the event in question. In other words, Meadow considered the one family in 8500 with a cot death to simply be unlucky, with no biological or environmental diathesis for the condition, rather like the probability that one will roll a six with a die. Thus, according to Meadow, the second death was no more likely than the first, just as a second six is no more likely than the first, and the probability of each can be multiplied to arrive at the probability of both.

The most obvious problem with this thinking was that it presumed that all possible biologic and environmental risk variables for cot death have been investigated and are known, and that none were present in the Clark family; an over simplification to the point of deception. The approach could be compared to the superficially convincing claim that an injury by lightning strike is a random event, and thus a second such injury would be no more likely than the first. In reality, an individual who works on a golf course in a geographic area with frequent thunderstorms is much more prone to a lightning strike injury than is an office worker who resides in an area where such storms are rare.

The guilty verdict was appealed, based in part on the problems with Meadow's statistical claims. The conviction eventually was overturned when it was found that a witness had failed to reveal evidence that one of the children's deaths may have been associated with a *Staphylococcus aureus* infection, among other problems.[6] The public attention brought to the Clark case also focused attention on the lack of validated criteria necessary for a diagnosis of MSbP, with the result that several murder convictions that had resulted from Meadow's testimony were reviewed.

The tragedy of the Clark case raises the question of how such a chain of events could have occurred; how could testimony that is fundamentally flawed be used to imprison an innocent person? An issue of even greater concern is how often similar superficially convincing testimony is used effectively in criminal and civil proceedings, resulting in an unjust verdict. What is apparent from the many different experts from a variety of disciplines who gave opinions regarding the faulty testimony in the Clark case, is that not only is there uncertainty regarding what constitutes valid testimony on some issues, but that there is also a degree of uncertainty as to who should be setting the standards for such testimony. As an example, while it is not difficult to determine that the relevant expert to testify in a medical malpractice claim against an oncologist whose patient died from hypovolemic shock while receiving autologous stem cell therapy is another oncologist, who is the appropriate expert to address a claim that the patient would have died within 5 years, on a more likely than not basis?

Any expert opinion that addresses the probability, risk, incidence, or prevalence of an event occurring or not occurring in an individual or a population is an opinion that must have a foundation in valid epidemiologic concepts and data. Epidemiology is most simply defined as the scientific study or analysis of populations having similar disease or injury characteristics. The proper application of epidemiologic concepts and data to forensic issues is the practice of Forensic Epidemiology. The term was first introduced by Loue in 1999[7] and later adopted by the US Centers for Disease Control and Prevention (CDC) in 2003 as a narrowly focused Public Health Law Program module designed to aid with the investigation of acts of bioterrorism.[8] The subject matter covered by the term Forensic Epidemiology has since been expanded to cover the multitude of areas in which epidemiologic terms, concepts, and data may be applied in a forensic venue.[9]

2. Forensic Epidemiology

The practical application of Forensic Epidemiology (FE) concerns both the recognition of the improper use of epidemiologic concepts and data as well as the use of such testimony to help prove the assertion of one side or another. For this reason, this paper is organized in three sections; the first will help the reader identify the most common scenarios in which experts (both epidemiologic and non-epidemiologic) use terms and concepts of epidemiology in forensic venues. The next section describes appropriate uses of epidemiologic concepts and data in forensic venues. The final part of the paper is devoted to common fallacies associated with epidemiologic and probabilistic testimony in forensic venues. Because population-based inferences are often-used to support variety of expert opinions, the tenets of FE, as presented in this paper, are directed at all experts who give opinions regarding medicolegal issues in forensic venues.

2.1. Common testimony types involving FE concepts

Probability serves as the basis for many decisions people make on a daily basis. When one purchases a new DVD player and opts not to also purchase an extended warranty, one is evaluating, consciously or subconsciously, a series of probabilities. Many factors – how much the unit costs, one's prior experience with DVD player failure, and how long one typically keeps electronics before replacing them with an updated model – affect the decision that it is less than likely that the expenditure on the warranty is justified. If one is given some additional knowledge, for example, that one out of every three DVD players will require a costly repair within a year after the manufacturer's warranty expires, then one can assess the risk one is taking in not purchasing the warranty.

Please cite this article in press as: Freeman MD et al. Forensic Epidemiology: A systematic approach to ... J Forensic Legal Med (2008), doi:10.1016/j.jflm.2007.12.009

In a similar fashion, data and terms of probability may be used to sway judge and jury fact finders in a forensic setting, because assigning a weight to an opinion is a common method of strengthening testimony. These opinions affect how a fact finder perceives issues such as causation, negligence, and injury severity and prognosis that dictate trial outcomes.

2.1.1. A "reasonable probability"

The use of probabilistic language is inescapable in the forensic setting, since the standard for admitting expert testimony is that it is rendered as being "more probable or likely than not" or as a "reasonable probability" or "reasonable medical probability"; all relatively interchangeable terms.[10] In some jurisdictions, this standard for expert opinion is assigned a value that must be exceeded before the testimony is admissible; the expert must be "more than 50% certain" that the opinion is correct. Using probabilistic language for such testimony is somewhat of a mischaracterization of an internal process of the expert, who has opined that he is more certain than not that his opinion is accurate or true, regardless of the methods used to arrive at the opinion. There is no way objectively to weigh all of the processes that make up such a standard, because experts potentially are influenced by many different factors that may cause them to favor a particular opinion. One example of an exception, in which the decision processes of the expert can be externally scrutinized, is when a data set is described and an opinion is rendered that a particular outcome lies inside or outside of an error range or confidence interval bracketing the average of the data set. In such an instance, an opinion that it is a reasonable probability that an outcome would *not* take place, for instance, the failure of a medical device, could be reached based upon the application of valid and relevant epidemiologic/statistical tenets to the relevant data.

Probabilistic opinions may be expressed in terms of the:

- *Incidence of an occurrence or condition.* This is expressed as a rate, with a number of affected persons per some denominator. For example, the rate of traffic crash fatalities in the United States in 2004 was 1.45 per 100,000,000 vehicle miles traveled, 14.59 per 100,000 persons, 18 per 100,000 registered vehicles, and 21.54 per 100,000 licensed drivers.[11] Although the numerator and denominator are different for each estimate they all represent the same total number of deaths for 2004.
- *Prevalence of an occurrence or condition.* Prevalence is essentially a cross-section of a population at a given point in time. It is expressed as a proportion or percentage, such as the prevalence of cancer in the United Kingdom is approximately 2%, meaning that one out of every 50 live persons has a diagnosis of cancer in the UK.[12] Prevalence can be estimated from the incidence and survival rate of a condition.
- *Risk of an occurrence or condition.* Typically given as a proportion or percentage, this is the most commonly used and abused epidemiologic concept in forensic testimony. Risk may be expressed in absolute terms, *e.g.* the risk of dying in a motor vehicle crash in a given year is approximately 1 in 6500,[13] or as a relative risk (typically presented as a ratio, but also as a difference), such as the lifetime risk of dying in a car crash is more than 23,000 times greater than dying from a snake bite.[3] Misunderstanding is rife in such claims, however. For example, while it is reasonable to conclude that one is significantly less likely to die from a snake bite than in a traffic collision, this does not mean that handling a venomous snake is safer than driving a car. The average person's exposure to a venomous snake, in terms of duration, may be more than 23,000 times less than their exposure to a motor vehicle; thus the incidence of snake bite death may be significantly higher per unit of time of exposure than that of motor vehicle death per exposure for the same unit of time.

Opinions involving risk often rely upon probabilistic language, and this in turn may lead to a lack of specificity. For example, it is reasonable to opine that not wearing a seatbelt increases the risk of ejection in the event of a rollover crash, an important determination in some forensic venues, as it may indicate contributory negligence of the occupant to his or her own injuries for failure to wear a seatbelt. On the other hand, if the occupant was not wearing a seatbelt, and was *not* ejected but still fatally injured, the presence or absence of a seatbelt is a significantly smaller factor for injury frequency and severity, particularly if there is a great deal of vehicle roof crush that may have resulted in severe head and neck injury to a properly positioned and restrained occupant.[14] Quantification of the difference in risk of injury between the two scenarios would be important in helping a fact finder (judge or jury) determine whether the lack of a seatbelt was a significant factor in the case in question.

The following are a sample of common forensic opinions that state or imply probability (*NB* – the validity of the opinions is not addressed):

- *It is more likely that the patient will need future surgery as a result of the injury*
  - This is a prediction of an event that has not yet occurred. It implies that the future *incidence* of surgery for those who have the injury is higher than for those who do not. Such claims do not necessarily have to be based upon published epidemiologic data, because they can be a statement of clinical experience, based on one or both of two observations: that patients with the injury in question go on to have the surgery more often than patients who do not, or that an disproportionately large percentage of patients who have the surgery have a history of the injury. Similar claims, of what has been observed

and thus is *possible* or *plausible,* are the converse of statements of what is *impossible*, with regard to the level of substantiating data needed for a valid conclusion. As an example, if one wanted to determine if any red headed subjects were included in a group of 100, a sample size of one could establish the fact, if a read head was included in the sample. Conversely, if one wanted to establish that there were no red heads in the group one would have to examine the hair color of every group member. The level of proof required to validate a claim of possibility or plausibility is significantly less than what is required to establish impossibility or implausibility.

- *If the occupant had worn his seatbelt the injury would not have occurred*
  - The statement implies that the *risk* of injury for similar crashes with similar occupants who are restrained is 0. Unlike the previous opinion, such a statement implies a basis in data gathered from large samples of crashes and occupants that are similar to the case in question, as it implies impossibility of an outcome not just within the realm of the expert's experience but for all restrained occupants exposed to the same type of collision.

- *The disc herniation was not caused by the fall because the patient did not have immediate acute pain, something that would have been expected with a traumatic disc herniation*
  - This is a statement of prevalence for a certain condition at a certain point in time; it refers to the status of all patients with a traumatic disc herniation shortly after the trauma has occurred. The claim forces an inference that 100% of such patients will have pain immediately following injury. It would be unusual, if not unheard of, to find a population sample that would allow for such a definitive and broadly sweeping inference. In some cases, however, there may be a physiologic reason for such a statement, as for example, when the presence or absence of evidence of hemorrhage is used to determine whether injury may have occurred pre or post-mortem.

- *Retinal hemorrhage in an infant is reliable indicator of shaken baby syndrome*
  - This claim sounds like an estimation of point prevalence, but in fact it is a statement of prevalence ratio (also known as odds) as it implies a comparison of the finding of retinal hemorrhage in violent assault versus some other trauma. Because many shaken baby syndrome homicide prosecutions are defended with the assertion that the injuries resulted from a fall or some other precipitating unintentional trauma, the claim implies that the incidence of retinal hemorrhage in infants that have sustained an unintentionally self-inflicted injury is very small – at or near zero, and that the incidence of retinal hemorrhage in infant victims of assault is significantly higher.[15] As a hypothetical example, it might be said that only 1% of infant fatalities that result from an unintentional injury result in retinal hemorrhage, whereas 75% of confirmed cases of shaken baby syndrome have the same finding. Thus, a finding of retinal hemorrhage is at least suggestive of homicide, absent any other evidence. What is less clear is how such evidence would be presented when there is a smaller difference between the two prevalence estimates, *e.g.* 30% versus 50%, and at what point the difference becomes insignificant from a evidentiary perspective. This is further discussed in Section 2.2.3.

2.2. Principles of applied Forensic Epidemiology

This section of the paper focuses on how FE is used to formulate or substantiate an opinion, and the principles governing such an application. FE is of little use in describing something that has already occurred and been observed; this is the job of the clinician. However, when there are questions regarding causation of injury, or multiple potential causes, or unknown outcomes, the probability that one cause played a greater role than another must be weighed, and this often requires the interpretation of data derived from epidemiologic study. An example would be a crash-related head injury associated with a multiple impact collision scenario, including both frontal and near side impacts, with the forensic question of which impact caused the injury. An FE approach would consist of evaluating the probability of a head injury for a near side impact in which the occupant's head is highly likely to sustain a high acceleration contact with an unyielding structure such as the B-pillar, in comparison with a frontal crash scenario in which peak head acceleration is typically lower. The basis for the opinion would have to come from analysis of real world data, as illustrated in the example in Fig. 1. The analysis may already exist in the literature or it may need to be conducted *de novo* for the purposes of the forensic investigation. The probabilistic data can not

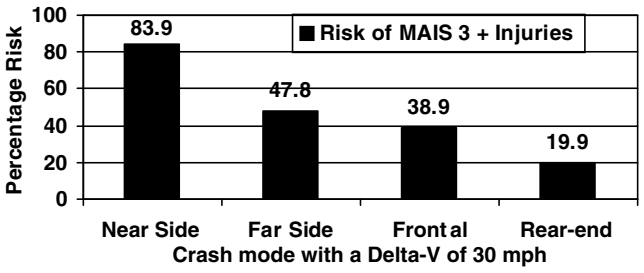

Fig. 1. Adapted from [Augenstein J, Perdeck E, Bowen J, Stratton J, Singer M, Horton T, Rao A. Injuries in near-side collisions. In: Proceedings of the 43rd annual conference of the association for advancement of automotive medicine; 1999. p. 139–58]. (MAIS 3+ refers to injuries rated as serious or greater on the Abbreviated Injury Scale.)

conflict with the real evidence; for example, if the medical evidence clearly showed an injury to the occupant's face then this would be given more weight that the probabilistic evidence. The FE conclusion should support, or be supported by, the evidence in the case.

Another application of FE is in the "what if" scenario in which an undisputed outcome is compared to a theoretical outcome if a predicate action or event had been different. For example, in a case of alleged medical negligence, in which an encapsulated ovarian granuloma tumor has been incompletely removed, the effect might be that the tumor is advanced from a FIGO Stage I (contained within the ovary) to a Stage II (spread to the pelvis). With adequate data, and based on the alternate "what if" scenario of the surgeon performing the procedure completely, the 5-year survival probability for the actual stage of the tumor (II) can be compared to that of the stage of the tumor had the alleged negligence not taken place (I).[16]

FE also has demonstrated utility in the criminal prosecution of alcohol-associated vehicular homicides in which some or all occupants were ejected, and there was dispute as to who was driving; the decedent or an ejected, as well as intoxicated, survivor.[17] This application of FE, called injury pattern analysis, evaluates injury patterns associated with various occupant positions, and relies upon probabilistic weighting of observed injury distribution and nature versus expected anatomic distribution and pattern of injury based upon observational epidemiologic study of similar types of collisions.[18]

2.2.1. Causation

Standards for epidemiologic determinations of cause and effect were first laid out in a systematic fashion by Hill in 1969.[19] Hill outlined nine criteria by which determinations of causation could be made when there is substantial epidemiologic evidence linking a disease or injury with an exposure, *e.g.* smoking and lung cancer. The criteria have since been modified and distilled by others[20–23] but they all comprise three basic elements:

1. There must be a biologically plausible link between the exposure and the outcome. Traumatic loading and bony fracture would be a straightforward example of a plausible link. An example of an implausible link would be trauma and leukemia. Plausibility is a very low threshold that can be overcome with relatively weak evidence, such as from small observational studies (case studies or case series with small numbers of subjects), or from the results of well-designed experiments with many subjects. Analogy also is a valid method of establishing plausibility; if forceful loading from one type of trauma can cause an injury, than forceful loading from another type of trauma may be a plausible cause of the same kind of injury.
2. There must be a temporal relationship between the exposure and the outcome. Quite obviously, the outcome cannot pre-exist the exposure. Less obviously, the outcome cannot postdate the exposure by a time period that is clinically considered to be too long or too short to relate the two. This determination is highly dependent upon the specifics of any case. For example, benzene exposure to the skin will cause symptoms of irritation that are apparent within 1 day, however, changes to the blood system may not be apparent for months. A crash-related injury to the spine may not be apparent for a day or two or even a week, but will not be completely latent for 2 months prior to causing symptoms. On the other hand, an injury that causes acute symptoms to the spine may mask or overlap with other symptoms resulting from, for example, a concomitant shoulder injury. Such an injury may not be apparent for some time following the original traumatic episode. The determination of etiology is typically made based on clinical judgment on a case-by-case basis, rather than from clearly delineated guidelines or principles.
3. There must not be any *likely* alternative explanations for the symptoms. The term "likely" is of critical importance, as, for example, it is not sufficient to simply point out that a patient with back pain following trauma is obese, that obesity is related to back pain, and thus it is more likely that the obesity rather than the trauma caused the back pain. For an alternative etiologic explanation to be considered more likely than an alleged exposure it must be both biologically plausible and have a stronger temporal relationship to symptom onset than the alleged exposure. If plausibility is present and temporality is relatively comparable, then two exposures can be compared by examining the dose–response of each exposure. This term typically refers to the magnitude and intensity of each exposure, but for the purposes of FE may also refer to outcome risk. An example of a comparison of dose–response is seen in the above example comparing the probability of head injury for a near side impact collision versus a frontal impact. Both are biologically plausible mechanisms of head injury and both occurred at the same time; however, the near side impact scenario has an established higher head injury risk.

It is common practice for clinicians, rather than epidemiologists, to make determinations of causation in individual patients. A clinician's causal determination incorporates the patient's history and the results of examinations and tests with the clinician's experience and training to arrive at a conclusion regarding causality. Such determinations, however, may not violate any of the three basic elements of causation.

2.2.2. Strength of evidence

In 1993, the United States Supreme Court issued an opinion in a case called Daubert v. Merrell Dow Pharmaceuticals Inc. This case set new standards for evidentiary hearings in the United States, in which the judge acts as a gatekeeper for proposed scientific testimony.[24] The case

concerned the alleged teratogenic effects of the drug Bendectin, used primarily for pregnancy-associated morning sickness. The plaintiff in the case had brought forth evidence from a variety of experts who cited *in vitro*, animal, and chemical studies as a basis for their collective opinion that Bendectin caused the birth defects that were the subject of the lawsuit. In response, the defense produced an epidemiologist expert who presented an analysis of epidemiologic (observational) studies of women who had used the drug, and opined that there was no relationship between the use of Bendectin and birth defects. A lower court had ruled that the experimental evidence presented by the plaintiff was insufficient to establish causation in light of the epidemiologic evidence of the defendant. When the case reached the Supreme Court the Justices ruled in favor of the defendant, affirming the ruling of the lower courts and establishing a new set of criteria for the admissibility of expert scientific testimony. The Daubert decision helped to highlight the use and misuse of forensic scientific evidence to establish or question causation. In this decision, a causal relationship suggested or refuted by an animal or cadaveric study is insufficient proof for establishing the etiology of an injury or disease when there is contradictory observational evidence. The latter includes clinical determinations of causation that do not violate the three basic elements of causation noted above. Fig. 2 illustrates the hierarchy of evidence strength in terms of its utility in establishing causality. The Daubert decision solely addresses evidence that is intended to support or refute the first element of causation, biologic plausibility. Temporality and likely alternative explanations are primarily clinical determinations and thus largely unaffected by Daubert.

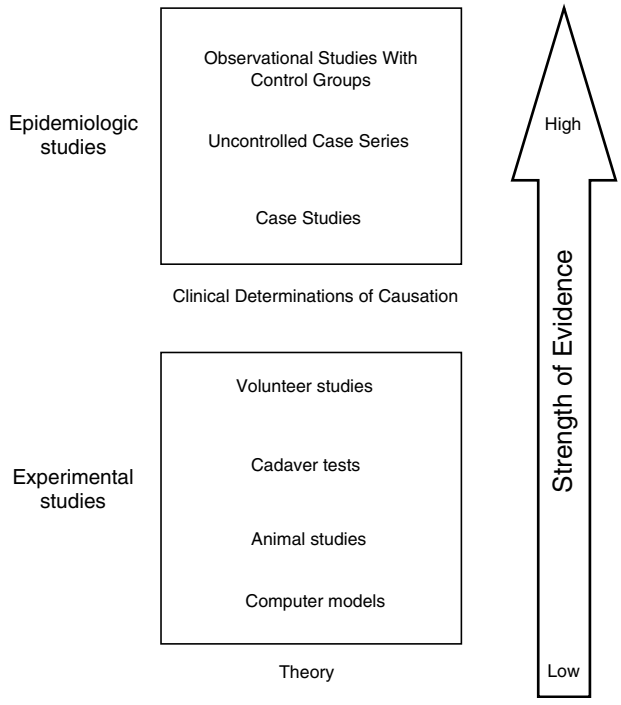

Fig. 2. Hierarchical depiction of evidence strength for causal determinations modeled from the Daubert decision.

2.2.3. Sensitivity, Specificity, and Positive Predictive Value

These are fundamental epidemiologic concepts that are critical to appropriate weighting of testing results. These concepts are equally important for understanding the precision or reliability of a particular opinion based upon the interpretation of a fact in evidence using an established set of criteria. For any test or criterion there are at least four possible results: true *positive* (TP), in which the test correctly identifies the guilty (or liable) party, true *negative* (TN), in which the test correctly identifies the innocent party, and false *positive* (FP) and false *negative* (FN), in which the test incorrectly identifies the innocent as guilty or the guilty as innocent, respectively. In a criminal forensic setting, the sensitivity of a test indicates the percentage of guilty defendants the test correctly identifies as guilty, and is calculated by dividing the true positives (the correctly identified guilty defendants) by all of the guilty defendants, included those incorrectly identified as innocent (TP + FN). The specificity of the test indicates the percentage of innocent defendants that the test will correctly identify, and is calculated by dividing all of the correctly identified innocent defendants (the true negatives) by all of the innocent defendants (TN + FP). The most important parameter of a test or criterion that may be used in a forensic venue is its Positive Predictive Value (also called Predictive Value Positive), as this value indicates how often the test is correct when it indicates guilt. As such, PPV measures the potential for harm when a particular test is used as an isolated index of guilt, as it also demonstrates the proportion of innocent defendants incorrectly identified as guilty. Table 1 is matrix that illustrates these measures.

For purposes of illustration, an example of the utility of Positive Predictive Value can be made with the claim made earlier in this paper, that retinal hemorrhage (RH) is a reliable indicator of shaken baby syndrome (SBS). For the following example it is assumed that the claim that RH is "reliable" equates to a sensitivity of 90% and a specificity of 75%; that is, RH is present in 90 out of 100 cases of SBS, and 75% cases of non-SBS death will not have RH. If one were to present these statistics as support for the opinion that a pediatric death resulted from SBS because RH was present it would likely be given a great deal of weight in a forensic setting. If the contrasting defense theory is that the fatal injury and RH resulted from a fall instead of a violent assault, the Positive Predictive Value (PPV) of RH as an indicator of SBS can help a jury determine the weight they should give to the evidence. In order to calculate PPV, however, it is necessary to know more about the data underlying the sensitivity and specificity calculations. If, for example, there are 200 cases of SBS deaths annually, and this results in 180 (90%) with findings of RH, and there are 1000 cases of non-SBS head injury annually, with only 250 (25%) with findings of RH, then the PPV is only 42% (see Table 2), and a determination of SBS based

Table 1
2 × 2 matrix illustrating the relationship between the Sensitivity, Specificity, and Positive Predictive Value of a test for guilt

		Criterion			
		+	–		
Guilt	Yes	True Guilty (TP)	False Innocent (FN)	All Guilty (TP + FP)	Positive Predictive Value TP/(TP + FP)
	No	False Guilty (FN)	True Innocent (TN)	All Innocent (FN + TN)	Negative Predictive Value TN/(FN + TN)
		All + tests (TP + FN) Sensitivity TP/(TP + FN)	All – tests (FP + TN) Specificity TN/(FP + TN)	All cases (TP + FP + FN + TN)	

TP and FP are True Positive and False Positive and TN and FN are True Negative and False Negative.

upon the finding of RH alone would be improper (*NB*: the figures used in the above example are solely for illustration).

2.3. Common forensic fallacies involving epidemiologic concepts

2.3.1. Prior Odds Fallacy

The Prior Odds Fallacy, described for the first time in this paper, is related to the Prosecutor's Fallacy, in which the pre-event or predictive odds associated with a piece of evidence are presented to the jury as a gauge of guilt.[25] An example of the Prosecutor's Fallacy is as follows: a rare blood type, present in only 1% of the population and matching that of a suspect is found at a crime scene. The prosecutor uses this evidence to suggest that there is a 99% probability that the suspect is guilty. The reason the inference is a fallacy is that while the matching blood type is *suggestive* of guilt, the 99% figure is unrelated to the *certainty* of guilt. For example, in a city of 1,000,000 there would be 10,000 people with the same blood type as the suspect. Or, the crime may have taken place in an ethnically homogenous community where 90% of the local denizens have the rare blood type.

In contrast with the Prosecutor's Fallacy, the Prior Odds Fallacy, typically offered in injury litigation settings as evidence against or for causality, is not suggestive of either. The Prior Odds Fallacy is seen when the low probability of an occurrence, *e.g.* possessing a winning lottery ticket, is used to cast doubt on the accuracy of the observation of the occurrence.

The following example illustrates the fallacy: a woman is involved in a rear impact collision that results in minimal damage to her vehicle, and is subsequently diagnosed with a permanent spine injury by her doctor. The insurer defending the case hires a doctor who examines the patient and opines that the majority of crash injuries recover spontaneously within a matter of months following a crash with minimal damage, and therefore it is highly improbable that the signs and symptoms of permanent injury are related to the collision in question. The Prior Odds Fallacy was committed in the example when the pre-crash or "prior odds" of contracting a permanent injury (say 1 in 20 or 0.05) was used suggest a correspondingly high probability (19 out of 20, or 0.95) that the original doctor's determination of causation was in error. The fallacy occurs due to the fact that there is no relationship between the probability of injury in the general population exposed to minimal damage crashes (0.05) and the frequency of clinician error in determining causality in patients that have been exposed to minimal damage crashes (unknown, but unlikely to be 0.95). The pre-event probability of an occurrence is not a valid measure of whether the occurrence took place; either it did or it did not (0.0 or 1.0). As a simple example, deaths resulting from plane crashes are exceedingly rare, however, a pathologist's clinical observations of a decedent following a plane crash would not be considered to be in error because the death was unlikely to have occurred.

The fallacy can be further illustrated with the example of the roll of a six-sided die. The probability that a six will be rolled is 1 in 6 (17%), and the probability that something other than a six will be rolled is 5 in 6 (83%). In this example,

Table 2
2 × 2 table illustrating the Positive Predictive Value of retinal hemorrhage presence as a gauge of guilt

		Retinal hemorrhage			
		Present	Absent		
SBS	Yes	True Positive (TP) 180	False Negative (FP) 20	All SBS + (TP + FP) 200	Positive Predictive Value TP/(TP + FN) **42**%
	No	False Positive (FN) 250	True Negative (TN) 750	All SBS – (FN + TN) 1000	Negative Predictive Value TN/(FP + TN) 25%
		All cases with RH (TP + FN) 430	All cases without RH (FP + TN) 770	All deaths (TP + FP + FN + TN) 1200	
		Sensitivity TP/(TP + FP) **90**%	Specificity TN/(FN + TN) **75**%	Prevalence of SBS in all deaths (TP + FP)/(TP + FP + FN + TN) 17%	

the result of the roll is recorded by a hypothetical machine that has been found to have an error rate of one in 50 observations, so that the roll is misidentified in 2% of cases. The Prior Odds Fallacy would occur if a 6 was rolled and subsequently identified as such by the machine (with a 2% error rate), but it was asserted that there was an 83% probability that the result was something other than a 6 (83% probability the machine is wrong).

The error rate of clinical determinations of causality in minimal damage crash injuries is not known, but it is not likely to be very large. Such determinations would depend upon the observation of the three criteria of causation described earlier in this paper. Conversely, the error rate resulting from the introduction of evidence that invokes the Prior Odds Fallacy, if a judge or jury determination is based upon such evidence, would be calculated as follows:

$$\text{Prior Odds Fallacy error rate} = 1 - E_c,$$

where E_c is the actual error rate in clinician observations of causation. Although there are no published data on such errors, for the purposes of this paper the error rate is assigned a value of 0.05, theoretically taking into account cases in which false patient or physician attribution of causation has occurred. Thus, using the values stated above, the rate error in fact finder determinations when the Prior Odds Fallacy is accepted as evidence is $1 - 0.05 = 0.95$ or 95%.

In criminal cases the only fact finder determination is the guilt or innocence of the accused. This scenario differs from civil litigation, in which the fact finder must determine (a) whether the defendant acted negligently, and if so then (b) whether the act of negligence is causally related to the alleged injuries, and if so, then (c) the amount of damages to be awarded. The Prior Odds Fallacy is primarily directed at the causation determination in civil litigation. Further, of the three causation elements identified earlier in this paper (biologic plausibility, temporality, and the lack of a likely alternative explanation), the Prior Odds Fallacy is directed mainly at biologic plausibility, relying upon the implication that a low prior odds (*e.g.*, only one in 20 will be injured) is an indicator of *implausibility*. In fact, a very low level of scientific or clinical evidence is required to assert a *plausible* biologic association between a noxious exposure and an injury outcome, as plausibility is either present or not, regardless of degree. Thus, assertions of low frequency of association between an exposure and an outcome are irrelevant to biologic plausibility.

The Prior Odds Fallacy also occurs in plaintiff expert testimony that intended to support causality in civil litigation. An example is sometimes seen in crash injury cases, in which photographs of extensive vehicle damage are used to elicit testimony that the degree and extent of injury observed in a patient are consistent with the crash, implying a low likelihood of error in the observation of causal association between the subject crash and the diagnosed injuries. Regardless of whether the frequency of injury is 5%, 10%, or 40% for similar crashes, diagnostic error is independent of pre-event probability of injury.

2.3.2. Fallacies contributing to lower prior odds estimates

All of the following fallacies are used to establish an erroneous pre-event probability that would then be applied to causation via the Prior Odds Fallacy:

(a) *Non-representative sample fallacy.* Results observed in one side of a population distribution curve cannot be used to argue thatthe other side does not exist or is rare. An example is seen in the human subject crash testing literature in which some authors have concluded that there is a crash speed injury threshold below which injury is unlikely or impossible in the general population[26,27] with the intent that the thresholds be applied in medicolegal settings as a litmus test for causation.[28] The results of testing of the hardiest members of the population who are not injured until they are exposed to high crash forces (arrow "A" in Fig. 3) cannot be used to exclude the existence of, or in any other way define the distribution of the members of the population who are injured when exposed to low crash forces (arrow "B" in Fig. 3). The use of volunteer crash tests to establish an injury threshold in the general population has been criticized as unscientific, as the studies underlying the proffered forensic opinion suffer from (1) inadequate study numbers of subjects, vehicles, and crash conditions (contributing to random variation), (2) non-representative study samples (crash test volunteers cannot be said to represent the full spectrum of the motoring public with regard to injury susceptibility), and (3) they are conducted under non-representative conditions (crash tests are designed to minimize participant injury risk).[29] The fallacy also occurs when *in vitro*, *ex vivo*, animal model, computer model, and other surrogates are used as a basis for establishing or questioning causation.

(b) *Appeal to statistical authority.* Juries are more likely to be convinced of the validity of testimony when it is supported with a reference to statistics or statistical

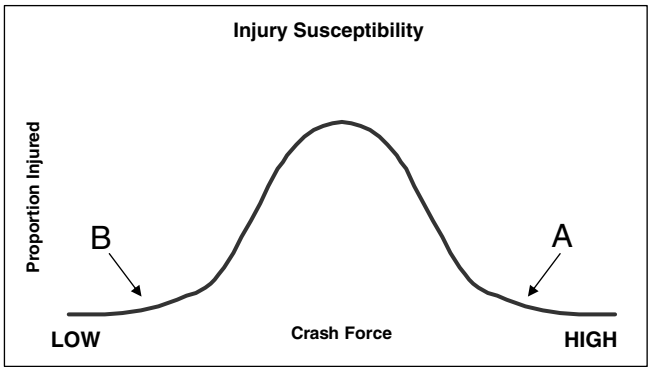

Fig. 3. Theoretical normal distribution of the general population by degree of crash severity required to cause injury.

language, regardless of the appropriateness of the statistical reference. An extreme example was seen in the Sally Clark case in which probability of innocence was given in very precise terms (1 in 73 million), however, non-specific probabilistic language tends to accomplish the same goal. The claim that an opinion is "highly likely" or "highly unlikely" rather than merely "likely" or "unlikely" is an example weighting an opinion without precision; unless the terms are defined and data are presented to substantiate the claims, such opinion weighting is speculative and potentially harmful. The same is true for the use of "reasonable certainty" versus "reasonable probability"; with no substantiating data the former is no more than an attempt to bolster the persuasiveness of an opinion with misleading language. Caution must be used when assessing for this fallacy; it does not occur when a clinician makes a claim based on clinical experience. For example, the claim that. "it is highly unlikely the surgery will result in a substantial impairment" is a reasonable conclusion for a surgeon to draw regarding one of his own patients, when the claim is based upon his experience performing the surgery and observation the results of the surgery. In contrast, it would not be appropriate for a clinician to claim that her clinical experience allows her to draw the conclusion that certain injuries are highly unlikely to require treatment, because many patients may leave this doctor's care and seek care elsewhere, unbeknownst to her. In the two examples above, the former claim is based upon a reasonable sample of the population to which the claim is to be extrapolated (patients who have surgery). The latter claim is less valid because the patients seen by any one clinician may be non-representative of the general injured patient population (*i.e.* specialist practices), and thus limit the extrapolability of the clinician's experiences.

(c) *Impossibility fallacy.* This occurs when an expert opines that a causal relationship is impossible; the claim implies both 100% clinical observation error rate (as a result of the Prior Odds Fallacy) and a nearly census of population-based data from which to infer the claim of 0 incidence. The fallacy also occurs when an expert claims that a causal relationship is always present, implying zero uncertainty in the opinion. The easiest way to understand this fallacy is to picture a box full of 100 rubber balls that can be either red or black. A statement that there is a red ball in the box (red ball possible) could be verified with a sample of only one ball, if the ball was red. In contrast, a statement that there are no red balls in the box (red ball impossible) would require an examination of every ball in the box. The claim may incorporate other fallacies as well, such as Non-representative sample fallacy, *e.g.* because no disc herniation has ever been observed in volunteer crash testing it is impossible to herniate a disc under similar circumstances.

(d) *Straw man fallacy.* This fallacy stems from the use of unvalidated constructs or inappropriate proxies for causal mechanisms. A good example is seen in the literature regarding whiplash injury; some authors have compared peak accelerations recorded at the head during activities such as sneezing[30] or skipping rope[31] to the accelerations observed in volunteer crash testing. The fallacy occurs when peak head acceleration is used as a proxy for injury risk, so that the improper conclusion is drawn that skipping rope and crash-related trauma have the same injury potential. Because whiplash injury occurrence is associated with numerous variables aside from peak acceleration, such as gender, occupant bracing, vehicle and seat variations, *inter alia*,[28] selecting a variable that is only loosely correlated with injury occurrence as an index of the probability of injury presence lies at the heart of this fallacy.

2.3.3. Base rate fallacy

This fallacy has been described by others as applying to a variety of scenarios but is frequently overlooked in forensic medicine testimony.[32] This fallacy occurs when the base rate of a finding or occurrence in a relevant comparison population is erroneously overlooked while the prevalence of the same finding in the target population is used as evidence in favor of one side or another. An example was presented previously in this paper, in the shaken baby syndrome (SBS) example. While it is important for the fact finder to be made aware that retinal hemorrhage is present in 90% of SBS cases at post-mortem examination, it is equally important to know the prevalence of the same condition in all of the relevant non-violent assault injury mechanisms as well, as discussed earlier in this paper in Section 2.2.3.

3. Conclusions

The 19th century essayist and novelist Charles Dudley Warner (1829–1900) is credited with the quote "Everyone complains about the weather but no one does anything about it". In some ways, the quote is apropos for the widespread but unsystematic use of probability in forensic medicine, in that everyone uses it but not everyone understands it. The purpose of this paper, in which the concept and some of the applications of Forensic Epidemiology have been introduced, is to fill a void that presently exists in forensic medicine with the addition of a general heading under which the proper and improper forensic use of probability is systematically described. As demonstrated by the tragedy of the Sally Clark case, there is little doubt that the use of probability in forensic medicine is in need of

Please cite this article in press as: Freeman MD et al. Forensic Epidemiology: A systematic approach to ... J Forensic Legal Med (2008), doi:10.1016/j.jflm.2007.12.009

standardization; there is a high potential for continued harm and injustice if nothing is done in this regard.

Better and more explicit heuristics are needed to describe and implement the concepts introduced in this paper for the wide variety of circumstances encountered in forensic medicine. A few recommendations are as follows:

- Be alert for the language of probability or epidemiology in forensic opinions.
- When epidemiologic data are referenced as a basis for an opinion, evaluate the propriety of their use. Are the sample population and circumstances sufficiently similar to allow for extrapolation to the facts in the present case?
- When in doubt regarding causal determinations, return to the three essential elements of causation: biologic plausibility, temporality, and lack of likely alternative explanation.
- When a clinical outcome is known, be aware of the potential for Prior Odds and other fallacies.
- If a test or criterion is set as an evidentiary standard, determine if the Specificity, Sensitivity, and Positive Predictive Value is known or can be determined for the test or criterion. Use these tools to help determine the real utility of the test or criterion in a forensic setting.

References

1. Bacon CJ. The case of Sally Clark. *J R Soc Med* 2003;**96**:105.
2. Meadow R. Munchausen syndrome by proxy. The hinterland of child abuse. *Lancet* 1977;**2**(8033):343–5.
3. Meadow R, editor. *ABC of child abuse*. London, UK: BMJ Publishing Group; 1989.
4. http://www.rss.org.uk/docs/Royal%20Statistical%20Society.doc [accessed 02-02-2007].
5. Watkins SJ. Conviction by mathematical error? *BMJ* 2000;**320**:2–3.
6. http://news.bbc.co.uk/1/hi/england/2708737.stm [accessed 02-02-2007].
7. Loue S. *Forensic Epidemiology: A comprehensive guide for legal and epidemiology professionals*. Carbondale: Southern Illinois University Press; 1999.
8. http://www2.cdc.gov/phlp/ForensicEpi/background.asp [accessed 11-27-2006].
9. Rossignol AM. *Principles and practice of epidemiology; an engaged approach*. New York, NY: The McGraw-Hill Companies; 2005. p. 224–5.
10. Mossman D. Interpreting clinical evidence of malingering: a Bayesian perspective. *J Am Acad Psychiatr Law* 2000;**28**(3):293–302.
11. http://www-fars.nhtsa.dot.gov/ [accessed 11-21-2006].
12. http://info.cancerresearchuk.org/cancerstats/incidence/prevalence/ [accessed 11-21-2006].
13. http://www.nsc.org/lrs/statinfo/odds.htm [accessed 11-21-2006].
14. Herbst B, Forrest S, Orton T, Meyer SE, Sances Jr A, Kumaresan S. The effect of roof strength on reducing occupant injury in rollovers. *Biomed Sci Instrum* 2005;**41**:97–103.
15. Wyszynski ME. Shaken baby syndrome: identification, intervention, and prevention. *Clin Excell Nurse Pract* 1999;**3**(5):262–7.
16. Petignat P, de Weck D, Goffin F, Vlastos G, Obrist R, Luthi JC. Long-term survival of patients with apparent early-stage (FIGO I–II) epithelial ovarian cancer: a population-based study. *Gynecol Obstet Invest* 2006;**63**(3):132–6 [Epub ahead of print].
17. Freeman MD, Nelson C. Injury pattern analysis as a means of driver identification in a vehicular homicide; a case study. *Forensic Examiner* 2004;**13**(1):24–8.
18. Freeman MD. Injury pattern analysis as a means of driver determination in a vehicular homicide investigation. In: Proceedings of 16th Nordic conference on forensic medicine; 2006. p. 38–9.
19. Hill AB. The environment and disease: association or causation? *Proc R Soc Med* 1965;**5**:295–300.
20. Stephens MD. The diagnosis of adverse medical events associated with drug treatment. *Adv Drug React Ac Poison Rev* 1987;**6**:1–35.
21. Miller FW, Hess HV, Clauw DJ, Hertzman PA, Pincus T, Silver RM, et al. Approaches for identifying and defining environmentally associated rheumatic disorders. *Arth Rheum* 2000;**43**(2):243–9.
22. McLean SA, William DA, Clauw DJ. Fibromyalgia after motor vehicle collision: evidence and implications. *Traffic Injury Prev* 2005;**6**:97–104.
23. Evans AS. Causation and disease: a chronological journey. The Thomas Parran Lecture. *Am J Epidemiol* 1978;**108**(4):249–58.
24. Daubert v. Merrell Dow Pharmaceuticals (92–102), 509 U.S. 579; 1993.
25. Thompson WC, Schumann EL. Interpretation of statistical evidence in criminal trials; the Prosecutor's Fallacy and the Defense Attorney's Fallacy. *Law Hum Behav* 1987;**11**(3):167–87.
26. McConnell WE, Howard RP, Poppel JV, et al. Human head and neck kinematic after low velocity rear-end impacts: understanding whiplash. In: Proceedings of the 39th stapp car crash conference, #952724; 1995. p. 215–38.
27. Szabo TJ, Welcher JB, Anderson RD, et al. Human occupant kinematic response to low speed rear-end impacts. SAE tech paper series 940532; 1994. p. 23–35.
28. Castro WHM, Schilgen M, Meyer S, Weber M, Peuker C, Wortler K. Do "whiplash injuries" occur in low-speed rear impacts? *Eur Spine J* 1997;**6**:366–75.
29. Freeman MD, Croft AC, Rossignol AM, Weaver DS, Reiser M. A review and methodologic critique of the literature refuting whiplash syndrome. *Spine* 1999;**24**(1):86–98.
30. Allen ME, Weir-Jones I, Motiuk DR, et al. Acceleration perturbations of daily living: a comparison to 'whiplash'. *Spine* 1994;**19**(11):1285–90.
31. Rosenbluth W, Hicks L. Evaluating low-speed rear-end impact severity and resultant occupant stress parameters. *J Forensic Sci* 1994;**39**(6):1393–424.
32. Fantino E. Behavior-analytic approaches to decision making. *Behav Process* 2004;**66**(3):279–88.

Please cite this article in press as: Freeman MD et al. Forensic Epidemiology: A systematic approach to ... J Forensic Legal Med (2008), doi:10.1016/j.jflm.2007.12.009