



The Inns of
Court College
of Advocacy

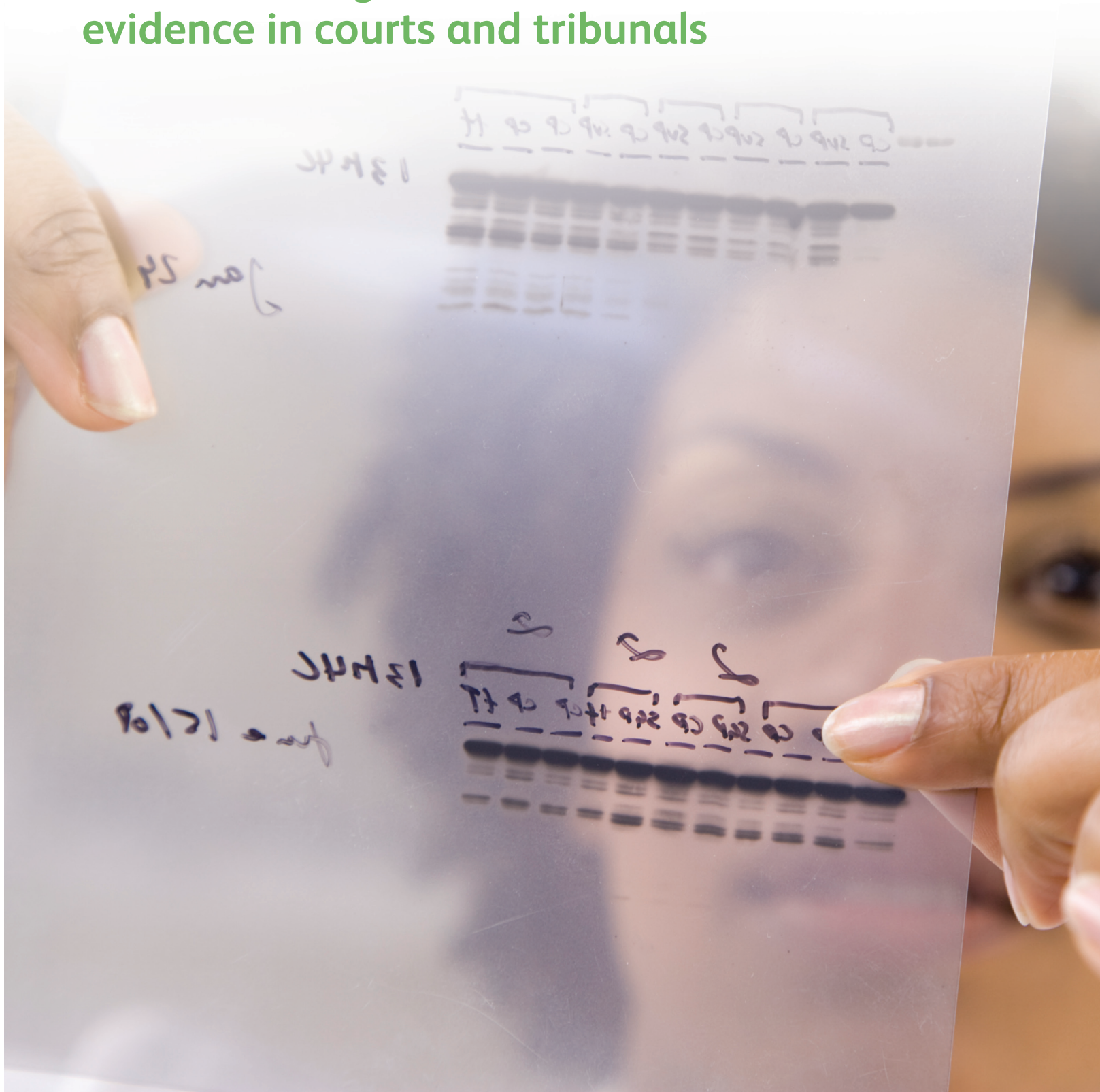
The Council of the Inns of Court



ROYAL
STATISTICAL
SOCIETY

DATA | EVIDENCE | DECISIONS

Statistics and probability for advocates: Understanding the use of statistical evidence in courts and tribunals



INTRODUCTION

An introductory guide for advocates produced by the Royal Statistical Society and the Inns of Court College of Advocacy as part of the ICCA's project: 'Promoting Reliability in the Submission and Handling of Expert Evidence'.

'The Inns of Court College of Advocacy is proud to have collaborated with the Royal Statistical Society over the production of this booklet. On the College's side it forms a major building block in the development of the training of advocates in the effective understanding, presentation and challenging of expert evidence in court. Experts in every type of discipline, appearing in every type of court and tribunal, habitually base their evidence on statistical data. A proper understanding of the way in which statistics can be used – and abused – is an essential tool for every advocate.'



Derek Wood CBE QC, Chair of the Governors of the ICCA

'More and more fields of expertise are using data. So expert evidence, whether in pre-trial or in court, will increasingly include a statistical element and it is vital that this is used effectively. The Royal Statistical Society started to work on statistics and the law following a number of court cases where the interpretation of statistics, particularly those presented by experts who were not professional statisticians, has been of concern. The RSS welcomes the Inns of Court College of Advocacy's programme to improve the reliability of expert evidence, and we hope this guide will ensure that evidence which includes statistics and data is used more effectively, for everyone's benefit.'



Sir David Spiegelhalter, President, Royal Statistical Society

CONTENTS

- INTRODUCTION..... 3**
- CONTENTS..... 4**
- HOW TO USE THIS GUIDE 5**
- STATISTICS AND PROBABILITY ARE IMPORTANT FOR ADVOCATES..... 7**
- REFRESHER: THE STATISTICIAN’S TOOLBOX..... 8**
- SECTION 1: STATISTICS AND PROBABILITY IN LAW..... 10**
 - 1.1 Misunderstandings of probability..... 10
 - 1.2 Coincidences and patterns 11
 - 1.3 Independence 13
 - 1.4 Absolute and relative risk..... 16
 - 1.5 Correlation and causation 18
 - 1.6 False positives and false negatives 24
 - 1.7 Prosecutor’s fallacy and defence fallacy 25
 - 1.8 Expert opinion evidence 29
- SECTION 2: KEY CONCEPTS IN STATISTICS AND PROBABILITY 36**
 - 2.1 Basic probability 36
 - 2.2 Inferential statistics and probability 38
 - 2.3 The scientific method..... 43
- SECTION 3: INTO PRACTICE 46**
 - 3.1 Communicating with experts about statistical evidence..... 46
 - 3.2 Guidelines for experts..... 51
 - 3.3 Case study one – Criminal 53
 - 3.4 Case study two – Civil (medical negligence)..... 55
 - 3.5 Case study three – Family..... 58
 - 3.6 Case study four – Planning..... 62
- SECTION 4: CURRENT AND FUTURE ISSUES 65**
- FURTHER RESOURCES..... 72**
- INDEX OF STATISTICAL TERMS 74**
- ACKNOWLEDGEMENTS 75**

HOW TO USE THIS GUIDE

This introductory guide will enable you more effectively to understand, recognise and manage statistics and probability statements used by expert witnesses, not just statistical witnesses. It also provides you with case studies which will enable you to think through how to challenge expert witnesses more effectively and assess the admissibility or reliability of their evidence.

Statistics and probability are complex topics. This booklet does not provide you with all the answers, but it aims to identify the main traps and pitfalls you are likely, as an advocate, to meet when handling statistical evidence. It will make you more aware of what to look out for, introduce you to the basics of statistics and probability, and provide some links to where you can find more information. We hope it will encourage you to investigate further opportunities for professional development in the understanding, interpretation and presentation of statistical and probabilistic evidence. We hope, too, that it will encourage you to consult more effectively with appropriate expert witnesses in the preparation of your cases.

A minority of advocates will have studied mathematics or science to a level that enables them instinctively to understand the detail of the statistical, scientific, medical or technical evidence which is being introduced in cases in which they are appearing. Most advocates need help, first in understanding that evidence, and then knowing how to deploy or challenge it. They must therefore be frank and open in asking for help.

The starting-point will usually be when the advocate meets his or her own expert in conference. Experience shows that statistical evidence is a particularly difficult topic to handle. This booklet will provide additional help in accomplishing that task.

The guide begins with a section which covers various examples of statistics and probability issues that may arise, or have arisen, in legal cases.

Section 2 covers basic statistics, probability, inferential statistics and the scientific method. You can use these brief chapters to brush up on your existing knowledge and fill in any gaps.

In Section 3, we provide advice on putting this all into practice, including guidelines for expert witnesses, and four case studies in different areas of law that will enable you to test your understanding, and consider how you might go about questioning expert witnesses. These case studies are intended to be realistic (though fictitious) examples that include several of the errors and issues referred to in this guide.

The fourth Section is a chapter on some controversial issues and potential future developments.

The booklet ends with a final chapter on further resources. The resources suggested here and throughout the guide will help you to develop your understanding and confidence further.

Shaded text such as this contains examples, explanations, and other material to illustrate the issues discussed. In section 3 it is used to distinguish between explanatory text and the case study text.

When first introduced, terms with specific statistical meaning are **highlighted** in bold and are included in the *Index of statistical terms*. Quotes are in *italics*. Italics are also used for *emphasis*. References to other sections within the document are *underlined and italicised*. Where possible, we have provided hyperlinks to documents that are available on the web; these appear green and underlined and can be accessed from the PDF version of this document.

STATISTICS AND PROBABILITY ARE IMPORTANT FOR ADVOCATES

Statistical evidence and probabilistic reasoning form part of expert witnesses' testimony in an increasingly wide range of litigation spanning criminal, civil and family cases as well as more specialist areas such as tax appeals, sports law, and discrimination claims. Many experts use statistics even though they may not be statisticians. It is important for an advocate to feel comfortable with any statistics and probability statements that might arise in any case.

One of the main reasons for having this knowledge is to avoid miscarriages of justice, some of which have occurred because of the inappropriate use of, and understanding of, statistics and probability by judges, expert witnesses and advocates. Had the advocates been able to cross-examine the expert witnesses more effectively, to expose the weakness of the opinion, these injustices might not have happened. In addition, addressing shortcomings at pre-trial stage may be even more effective, precluding the introduction of the evidence of the expert at trial.

This issue is being noticed by the media. The economist and journalist, Tim Harford, set out the challenges for advocates in a 2015 article for the *Financial Times*, 'Making a Lottery out of the Law'. He ends by saying:

Of course, it is the controversial cases that grab everyone's attention, so it is difficult to know whether statistical blunders in the courtroom are commonplace or rare, and whether they are decisive or merely part of the cut and thrust of legal argument. But I have some confidence in the following statement: a little bit of statistical education for the legal profession would go a long way.

The legal profession is aware of the importance of training in the understanding of statistics and probability. The Law Commission identified one of the main challenges faced by advocates as follows:

cross-examining advocates tend not to probe, test or challenge the underlying basis of an expert's opinion evidence but instead adopt the simpler approach of trying to undermine the expert's credibility. Of course, an advocate may cross-examine as to credit in this way for sound tactical reasons; but it may be that advocates do not feel confident or equipped to challenge the material underpinning expert opinion evidence.

References

Tim Harford '[Making a Lottery out of the Law](#)', *Financial Times* (London, 30 January 2015)

Law Commission, [Expert Evidence in Criminal Proceedings in England and Wales](#) (Law Com No. 325, 2011) para 1.21

REFRESHER: THE STATISTICIAN'S TOOLBOX

Statistics provides a set of tools to address questions and answer problems. These tools are typically applied in a process of **collecting** and **analysing data** and then **presenting** and **interpreting** information.

Here's a reminder of some basic principles and key terms for describing information and measurements. See also [Section 2: Key concepts in statistics and probability](#) and resources to improve your understanding that are listed in [Further resources](#).

Data is information about a subject of interest (for example, heights in a population) which can come in different forms. It can be quantitative (numerical) or qualitative (descriptive) – for example, the answers to interview questions. The important point about data is that for it to be of value to your case the data needs to reflect what you need to know. For example, data on the number of people with ponytails will not help you determine the number of people with 'long' hair. It is not a very good indicator of what amounts to or constitutes long hair or how prevalent it is. You would need to find a stronger measure of long hair such as 'hair over a predefined length measured from a particular point on the head'.

Ideally, if you want to know, say, the average height of a person in London, not only would you want to be sure how effectively the measuring was done, you would also want to collect information or data on everyone. Collecting data on everyone happens, for example, when we conduct a census. We can then make statements directly about a **population** - for example, its average age.

Generally, we can only **estimate** a particular characteristic or variable in the entire population because it is impractical or impossible to collect information on every instance. Rather we take a **sample** of the population we want to know about. We also want the sample to reflect the diversity of that population. In other words, the sample needs to be **representative**. There are techniques for ensuring this, and random sampling is important within this. See [2.2 Inferential statistics and probability](#).

Descriptive statistics

Descriptive (or summary) statistics illustrate different characteristics (or parameters) of a particular sample or population. It is useful to be familiar with some simple summary statistics of numerical data, particularly those that measure the 'central tendency' or average value, and those that measure how dispersed those data are.

The **mean** (or arithmetic mean) is the most commonly used 'average'. It is calculated by adding up the value of all the numbers reflecting a particular variable, and then dividing that sum by the total of all the numbers. For example, the mean of heights 68, 69, 71, 73, 73, 74 and 75 inches is 71.9 inches. The mean of hourly pay rates £7.85, £8.14, £12.69, £17.02, £18.21, £21.82 and £85.75 is £24.50. The mean is a sensible summary for data sets which are roughly symmetrical (have a similar number of data points above and below the mean), like height, but unlike pay.

The **standard deviation (SD)** is the most common measure of the spread or variability of a dataset. It provides a measure of how far, on average, different variables are from the mean. It provides a measure of relatively how much the data clusters around the mean, or whether the values are widely dispersed. The larger the standard deviation, the more spread out the data. The calculation of standard deviation is different for data from a population than from a sample. The sample standard deviation of the heights in the example given above is 2.6 inches.

The **range** is the difference between the highest and lowest value in any distribution, and can be affected by some extreme values (or **outliers**). The range for hourly pay in the example given above is the difference between £7.85 and £85.75, namely £77.90. Giving the minimum and maximum values to describe the range adds clarity.

The **median** is the middle number of the list which is created when you arrange numbers (or values) in order. When data are **skew**, that is, the data are not evenly distributed around the mean, such as the pay example in which 6 of 7 values are less than the mean, the median, £17.02, is a useful figure to use to describe the data; it is more typical. (If a list has an even number of values, then the median is calculated as being midway between the two middle numbers.)

The **interquartile range (IQR)** is an estimate of the middle 50% of the data. It is a sensible measure to use with a median, with both quartiles also stated, and is a more stable measure of spread than the range, since it excludes the extremes. For hourly pay, the IQR is the difference between £10.42, the lower quartile, and £20.02, the upper quartile - that is, £9.60.

The **mode** is the most frequent number. If seven employees take 0, 0, 0, 0, 1, 2 and 33 days of sick leave in a year, the mode is 0 days. The **frequency** of 0 days is 4, or 57% of the sample.

SECTION 1: STATISTICS AND PROBABILITY IN LAW

In this section, we look at examples of statistics and probability issues that may arise, or have arisen, in legal cases.

1.1 Misunderstandings of probability

The misuse of basic probability concepts (such as those in [2.1 Basic probability](#)) can lead to errors of judgement and bad decision-making.

Subjective probabilities

In 1964, a woman was pushed to the ground in Los Angeles, and her purse snatched. A witness reported a blond-haired woman with a ponytail, and wearing dark clothing, running away. The witness said she got into a yellow car driven by a black man with a beard and moustache. In this famous case, a couple were convicted of the crime.

The prosecutor’s expert witness, a maths professor, set out the respective probabilities for each of the characteristics and, multiplying each of those respective probabilities together, asserted a probability of 1 in 12,000,000 of any couple at random having them all.

Party yellow automobile	1 in 10	Girl with blond hair	1 in 3
Man with moustache	1 in 4	Black man with beard	1 in 10
Girl with ponytail	1 in 10	Interracial couple in car	1 in 1000

Koehler (1997) shows that this result was overturned on appeal on four grounds. Two of these reasons are dealt with in [1.3 Independence](#) and [1.7 Prosecutor’s fallacy and defence fallacy](#). The other two were first, that these were subjective estimates; and secondly that they did not account for the possibility that either or both of the suspects were in disguise. The air of certainty created by using numbers and mathematical formulae can sometimes blind people to ask searching questions.

Probability of the same thing happening again

Even if a fair coin comes up heads three times in a row, the probability that we get a head the next time is unaffected. It is still 50/50. The previous coin throws do not affect the outcome of subsequent coin throws. Even such basic misunderstandings of probability can occur in expert presentation of evidence, and may be prejudicial. See [1.3 Independence](#) for more.

References

CGG Aitken, 'Interpretation of evidence and sample size determination' in Joseph L Gastwirth (ed), *Statistical Science in the Courtroom* (Springer-Verlag 2000) 1–24.

Jonathan J Koehler, 'One in millions, billions and trillions: lessons from *People v Collins* (1968) for *People v Simpson* (1995)', (1997) 47 *Journal of Legal Education* 214

Ross v State (1992) B14-9-00659, Tex. App. 13 Feb 1992

1.2 Coincidences and patterns

We all tend to create patterns out of random information, or believe things are connected if they occur simultaneously. In legal cases, examples of miscarriages of justice have arisen from these 'cognitive biases'. Sometimes what looks like a pattern is just a coincidence.

A well-known example of an accusation of guilt being built up from coincidences, is that of Lucia de Berk, a Dutch paediatric nurse. It was alleged that her shift patterns at work coincided with the occurrences of deaths and unexpected resuscitations, and that the probability of this occurring by chance was 1 in 324 million. She was sentenced to life imprisonment in 2003 for four murders and three attempted murders. The High Court of The Hague heard an appeal in 2004 where it upheld the convictions and also found her guilty of three additional counts of murder, bringing the total to seven murders and three attempted murders.

At the initial trial, evidence was produced that in two cases she had administered poison to the victims. Dutch case law allows for 'chain evidence' to be used, where evidence of one offence can be used to support the case for a similar offence. In other words, if some murders have been proved beyond reasonable doubt (in this case by poison) weaker evidence is sufficient to link other murders.¹

The expert witness who calculated the 1 in 324 million figure had not used the correct method to bring together the probabilities of each separate incident, had incorrect data on de Berk's shift patterns, and had made assumptions that may not be reasonable – for example, assuming that all shift patterns are equally likely to have an incident occurring whether in summer or winter, day or night (Meester *et al.*, 2006). Making allowances for natural variations (heterogeneity), statisticians Richard Gill (who pushed for a retrial), Piet Groeneboom and Peter de Jong calculated an alternative chance of 1 in 26 that an innocent nurse could experience the same number of incidents as Lucia de Berk (Gill *et al.*, 2010).

The Commission for the Evaluation of Closed Criminal Cases (Commissie Evaluatie Afgesloten Strafzaken – CEAS) reviewed the case and reported to the Public Prosecution Service (Openbaar Ministerie) in 2007.

¹ In English law, the admissibility of previous behaviour or alleged behaviour as evidence relevant to the defendant's propensity to commit the offence charged is governed by section 101 of the Criminal Justice Act 2003.

They concluded that the case had been affected by ‘tunnel vision’ (in other words, seeking confirmatory evidence by initially focusing only on deaths that could involve Lucia de Berk, and not sufficiently considering alternative scenarios), and cast doubt on the evidence of poisoning of one child, suggesting that test results had been misinterpreted.

As a result of the review, the case was reopened, with further investigation and a retrial. It was subsequently shown that both alleged cases of poisoning were not supported, and de Berk was not on the premises for two of the alleged murders. Apparently, the overall number of deaths at that particular medical unit was also higher before she started working there. In April 2010, the appeals court in Arnhem ruled that there was no evidence that Lucia de Berk had committed a crime in any of the 10 cases, and she was acquitted.

Three plane crashes happened in an eight-day period in July 2014. Do you think this is suspicious?

Unpicking whether a set of events is a coincidence or not is not straightforward. We can start by considering the probability of the set of events occurring naturally (i.e. coincidentally), and then consider the probabilities of other explanations.

For the plane crashes, David Spiegelhalter, Professor for the Public Understanding of Risk at the University of Cambridge, has calculated that the chance of at least three crashes in an eight-day window is very small - around 1 in 1000 (Spiegelhalter 2014). But the right question to ask is whether such a ‘cluster’ is surprising over a longer period of time. He found that the probability of such a cluster of plane crashes within a 10-year period was about 60 per cent – i.e. more likely than not. It is not sufficient simply to identify this cluster to say that it is a suspicious pattern, rather than simply a chance event.

A 60 Second Science podcast ‘[Brain Seeks Patterns where None Exist](#)’, in *Scientific American*, in 2008, shows that people tend to see non-existent patterns. However, this ‘cognitive bias’ seems to be reduced if you are relaxed.

Kim Rossmo wrote in *The Police Chief* that: “Efforts to solve a crime by ‘working backwards’ (from the suspect to the crime, rather than from the crime to the suspect) are susceptible to errors of coincidence. These types of errors are often seen in the ‘solutions’ to such famous cases as *Jack the Ripper*.” The Lucia de Berk example also shows our tendency to link things together and create causal stories, as well as the need to carefully question when others might be doing the same.

It’s also important to consider the difference between the probability of a specific coincidence occurring, as opposed to the probability of any coincidence occurring within a group of events.

The probability of two people in a room of 30 sharing a birthday is approximately 0.71 - i.e. just over a 7 in 10 chance of it happening. In other words, it is more likely to happen than not.

If you are one of the 30 people in the room, the probability of any of the other 29 having the same birthday as *you* is approximately 0.08. That's less than one in ten chance of it happening.² In the first case, we are calculating the probability of a coincidence of the type of 'one person and another person share a birthday', in a group where there are many possible birthday pairings. In the second case, we are looking for a specific case of a pair where one of the pair is you.

A statistician can provide information to help the court understand the circumstances of the situation as it relates to the prosecution proposition and the defence proposition. It is not for the statistician to rule whether a set of events is a coincidence or not, guilt or not; that is for the court.

References and further reading

Richard D Gill, Piet Groeneboom and Peter de Jong, '[Elementary Statistics on Trial \(the case of Lucia de Berk\)](#)' (Richard D Gill, 2010) <www.math.leidenuniv.nl/~gill/hetero6.pdf> accessed 5 October 2017

Ronald Meester, Marieke Collins, Richard Gill, Michiel van Lambalgen, "On the (ab)use of statistics in the legal case against the nurse Lucia de B." (2006) 5 Law, Probability and Risk 233

Openbaar Ministeriem, 'Rapport Commissie evaluatie afgesloten strafzaken inzake mevrouw de B.', 29 October 2007.

D. Kim Rossmo, '[Failures in Criminal Investigation](#)' *The Police Chief* (October 2009) 54

Scientific American, '[Brain Seeks Patterns where None Exist](#)' (*Scientific American*, 3 October 2008) <www.scientificamerican.com/podcast/episode/brain-seeks-patterns-where-none-exi-08-10-03/> accessed 5 October 2017

David Spiegelhalter, '[Another tragic cluster - but how surprised should we be?](#)' (*Understanding Uncertainty*, 25 July 2014) <<https://understandinguncertainty.org/another-tragic-cluster-how-surprised-should-we-be>> accessed 5 October 2017

1.3 Independence

If the occurrence of an event has no effect on the probability of another event then the two events are said to be **independent**. There is no relationship between them. Where the outcomes are independent you can multiply the associated probabilities together to get the probability of the two outcomes occurring. But if the outcomes are related or dependent you cannot do so. Note that the relationship may not be obvious. Many pairs of outcomes that look independent are not (see [2.1 Basic probability](#) for a further explanation).

² For a full explanation of the mathematics of this, see <https://www.mathsisfun.com/data/probability-shared-birthday.html> for an illustrated example

A **variable** is a statistical term that refers to a quantity that could in principle vary or differ – for example, height of one person over different times of the day, or height of different people across a population. An **event** is an outcome that can either happen or not. Variables and events can both have a quality called independence.

In the case of the theft of the purse in Los Angeles described in [1.1 Misunderstandings of probability](#), the expert witness multiplied individual probabilities together to get 1 in 12,000,000 as the probability of any couple at random having all of six characteristics, such as “girl with blond hair” and “man with moustache”. Some of the stated identifiers are related and so cannot be multiplied together. In overturning the conviction, the Californian Supreme Court pointed out that since many black men have both beards and moustaches (including the defendant), there was “no basis” for multiplying the probability values for black men with beards and men with moustaches as these are not independent (*People v Collins*, 1968).

Sally Clark, a British solicitor, was convicted in 1999 of murdering two of her sons. One died shortly after birth in 1996, and another in 1998. The prosecution case included flawed statistical evidence given by an expert witness paediatrician who asserted that the chance of two children from a professional non-smoking family dying from Sudden Infant Death Syndrome (SIDS) was 1 in 73 million.

The figure of 1 in 73 million relied on a report which concluded that the possibility of a single SIDS in a similar family to Clark’s was 1 in 8,453. To reflect the two deaths, this probability was then squared to reach 1 in 73 million. The witness had presumed that each death was **independent** of the other and therefore multiplied the associated fractional probabilities of each death. The assumption of **independence** is not correct since there was evidence available that other factors, genetic or environmental, could have affected both children.

A second appeal in 2003 overturned the conviction when it was found that there had been a failure to disclose medical reports suggesting that the second son had died of natural causes. After her release, Sally Clark developed severe psychiatric problems and died from alcohol poisoning in 2007.

As the RSS said about the case of Sally Clark: “*The calculation leading to 1 in 73 million is invalid. It would only be valid if SIDS cases arose independently within families, an assumption that would need to be justified empirically.*” (Green 2002). In general, “*independence must be demonstrated and verified before the product rule for independent events can safely be applied*” (Aitken et al. 2010).

The Law Commission (2011) considered the Sally Clark case when recommending reforms for managing expert witnesses and said that if its reforms had been in place: “*the trial judge would have ruled on the scope of the paediatrician’s competence to give expert evidence and would*

have monitored his evidence to ensure that he did not drift into other areas.” It added: “The paediatrician would not have been asked questions in the witness box on matters beyond his competence; and if he was inadvertently asked such a question ... the judge would have intervened to prevent an impermissible opinion being given”.

Additionally, if the Commission’s reforms had been in place, the “defence or court would presumably have raised the matter as a preliminary issue in the pre-trial proceedings and the judge would no doubt have directed that the parties and their experts attend a pre-trial hearing to assess the reliability of the figure... The reliability of the hypothesis (or assumption) ... would then have been examined against our proposed statutory test, examples and factors. The expert would have been required to demonstrate the evidentiary reliability (the scientific validity) of his hypothesis and the chain of reasoning leading to his opinion, with reference to properly conducted scientific research and an explanation of the limitations in the research findings and the margins of uncertainty ...”

The Law Commission believed that, had this approach been followed, it “*would probably have provided a distinct basis upon which to quash C’s [Clark’s] convictions*” and that if a competent statistician had provided a more “*reliable figure as to the probability of two SIDS deaths in one family ... under our recommendations that expert would have been expected to try to formulate a counterbalancing probability reflecting the defence case.*”

Note that the Law Commission’s reforms have not been implemented in the form proposed by the Commission. See [Section 4: Current and future issues](#) for discussion.

References

Colin Aitken, Paul Roberts, Graham Jackson, [Fundamentals of probability and statistical evidence in criminal proceedings](#) (Royal Statistical Society 2010)

Peter Green, ‘[Letter from the President to the Lord Chancellor regarding the use of statistical evidence in court cases](#)’ (Royal Statistical Society, 23 January 2002) <www.rss.org.uk/Images/PDF/influencing-change/rss-use-statistical-evidence-court-cases-2002.pdf> accessed 5 October 2017

Law Commission, [Expert Evidence in Criminal Proceedings in England and Wales](#) (Law Com No. 325, 2011)

People v. Collins, 68 Cal.2d 319. [Crim. No. 11176. In Bank. Mar. 11, 1968.]

R v Clark [2000] EWCA Crim 54; [2003] EWCA Crim 1020.

1.4 Absolute and relative risk

Risk has different meanings or interpretations, depending on the legal, professional or social context. In finance, for example, risk may be shorthand for potential financial loss, or financial volatility. In everyday language, **risk** usually refers to the possibility of an outcome that is negative. In statistics, risk usually refers to the probability that an event will occur. This may be a good or a bad event. In epidemiology, risk usually denotes a measure of disease frequency (see definition of relative risk below).

In this guide, **risk** refers to the probability that an event will occur. In law, that event will constitute the harmful outcome that forms the basis of the legal claim. Two concepts of risk that often arise for consideration in tort litigation are relative risk and absolute risk, with the former being much more common.

Relative risk (RR) describes the proportional increase in the probability of an effect of an event occurring to a group, as measured from a baseline of a comparison group that has not experienced the event. The important point to take from this definition is that a statistical result can only properly be expressed in the form of a relative risk if it derives from a study in which one group with the relevant characteristic is compared with a control group over a specified period of time. The characteristic being studied may relate, for example, to the onset of a particular disease or to exposure to a harmful agent such as asbestos. That the 'RR' label can only attach to results stemming from rigorously implemented studies of this nature is a point that cannot be stressed enough. It should be borne in mind when looking at recent toxic tort cases. Questions to consider include: what is the source of the statistical figures being relied upon as indicators of factual causation? Is there a scientific basis for them? Are the studies relied upon relevant to the claimant in terms of age, gender and lifestyle? What generalizations have been made and are they tenable?

Absolute risk (AR) describes the overall risk to a population (or any other category) of an event – the probability of the event happening to any one member of the population. As such, it is not a term of art in the same way as relative risk. Indeed, it will often be the case that this statistic will have been extrapolated from an epidemiologic RR study. When assessing the relevance and reliability of such a statistic as regards the legal issue at stake, it will be of vital importance to look to the source of the information on which the AR calculation has been based. For instance, if the study in question was designed to measure the prevalence of lung cancer in women smokers aged between 25 to 40 years, it will not be relevant to a claimant who is male and/or a non-smoker. The smaller the study, the harder it will be to draw generalisations from it. In terms of assessing the scientific reliability of the data, obvious factors to look at will include the size of the P value (see

[2.3 The scientific method](#)) or confidence interval (see [2.2 Inferential statistics and probability](#)), and the measures taken to guard against biases (for example, bias in the selection of cases for a study) and confounders (see the ice cream and sunglasses example in [1.5 Correlation and causation](#) below).

Consider the following example. If you read: “People who use sunbeds are 20 % more likely to develop malignant melanoma”, this relates to the relative risk of developing cancer (assuming it derives from a reliable epidemiologic study with the all-important control group). This percentage on its own does not tell us what the overall likelihood is of developing this kind of cancer – in other words, the absolute risk.

Cancer Research UK has usefully explored the importance of understanding these differences. Williams (2013) noted that *The Guardian* reported a 70 % increase in thyroid cancer amongst women after the Fukushima nuclear disaster in 2011. This is a relative risk. A 70 % increase sounds alarming – but it does not tell us what proportion of the population would be affected. The *Wall Street Journal* reported that the absolute risk of this cancer was low (about 0.77 % - that’s seventy-seven out of ten thousand people). A 70 % increase in risk would take it to an incidence of 1.29 per cent in the population. That increase in absolute risk is 0.52 percentage points. For a population of 10,000, an extra 52 people would be affected. Expressing these changes in absolute risk allows for fair comparisons between different activities. For example, if another (fictitious) cancer had increased by 35 % after Fukushima, but had an existing absolute risk of 2 %, then the second cancer would have an extra 70 cases in 10,000 (from 200 per 10,000 to 270 per 10,000). In this example, although the change in relative risk was smaller (35 % rather than 70 %), the overall effect on the population was larger (an extra 70 cancer cases compared with an extra 52 cancer cases).

The important point to take away from the above example is that the same raw data can be used to make a number of different statistical calculations and so, when you are looking at expert evidence involving statistical results, you will need to know what calculations have been made and be able to explain the specific nature of the risk being presented. You will need to be able to determine at the case management stage the relevance and significance, or lack thereof, of the calculations to the specific legal issues you are dealing with.

References and further reading

Choudhury v South Central Ambulance Service NHS [2015] EWHC 1311 (QB) (an example of incorrect use of the technical term ‘relative risk’.)

Sarah Williams, ‘[Absolute versus relative risk – making sense of media stories](http://scienceblog.cancerresearchuk.org/2013/03/15/absolute-versus-relative-risk-making-sense-of-media-stories/)’, (Cancer Research UK, 15 March 2013) <<http://scienceblog.cancerresearchuk.org/2013/03/15/absolute-versus-relative-risk-making-sense-of-media-stories/>> accessed 5 October 2017

1.5 Correlation and causation

Complex phenomena with multiple potential relationships and impacts are sometimes explored by looking at **correlations** between variables rather than conducting an experiment (see [2.3 The scientific method](#)). This ‘observational’ method is used quite often in medicine to observe or indicate a link between various variables. It does not ‘prove’ causation but can be highly indicative of a relationship and lead to further research to determine causal mechanisms.

In 1965, an expert witness told a US Senate committee hearing that even though smoking might be correlated with lung cancer, any causal relationship was unproven as well as implausible. This followed the 1964 Report on Smoking and Health of the Surgeon General, which found correlations between smoking and chronic bronchitis, emphysema, and coronary heart disease.³

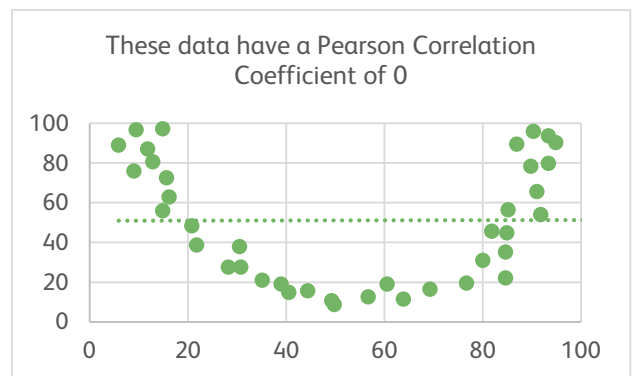
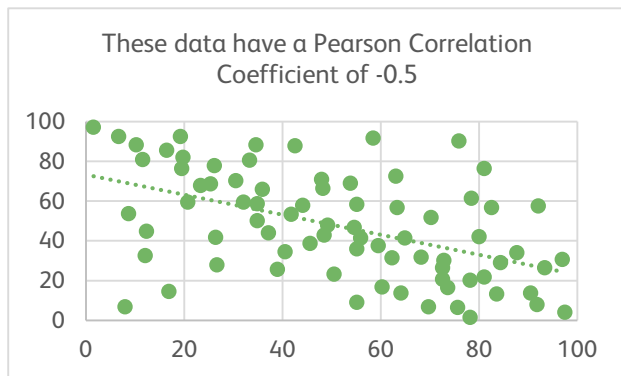
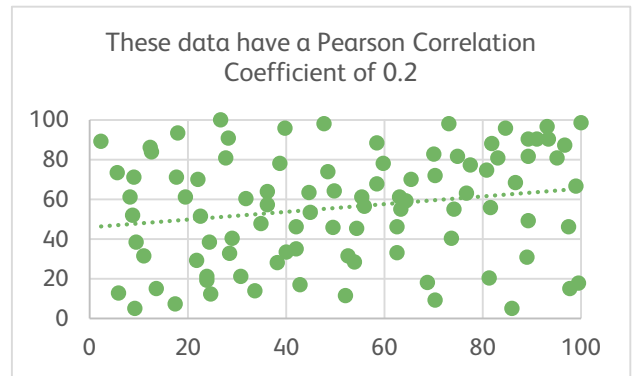
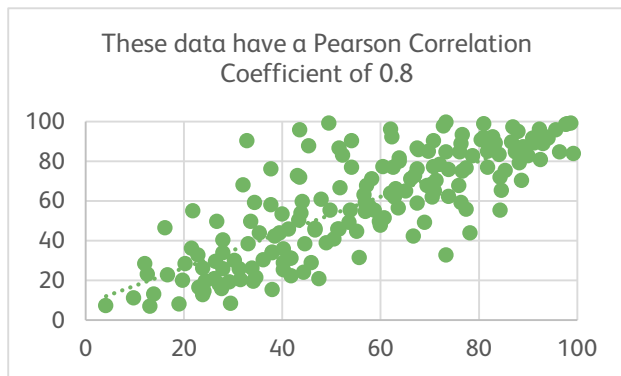
This example might at first glance be seen to illustrate the mantra that – ‘**Correlation does not equal causation**’. However, it is important to realise that many medical studies are, in fact, based on correlations between variables since the underlying causal mechanisms are not yet fully understood. These are called **epidemiological studies**.

The expert witness at the US Senate committee hearing on smoking, referred to above, was Darrel Huff. An esteemed populariser of statistics (see [Further resources](#)), he was *not* a qualified statistician. His testimony, and an unpublished book on the case for smoking, made statistical mistakes and did not look at the whole view (Reinhart 2014).

Correlations can be measured by specific statistics which calculate the likely extent of the ‘relationship’ between variables. An example is the Pearson Correlation Coefficient which calculates a number between -1 and +1 to describe the strength and form of the linear correlation. For example, +0.9 indicates a high positive correlation – as one value goes up, so does the other. If the result is 0, there is likely to be no linear relationship (although there may be another kind of relationship). If the coefficient is -0.5 there is a low negative correlation – as one value increases, the other decreases.

The Pearson Correlation Coefficient works only for linear relationships. There might be a curved pattern or relationship which would not be picked up. The charts overleaf illustrate this:

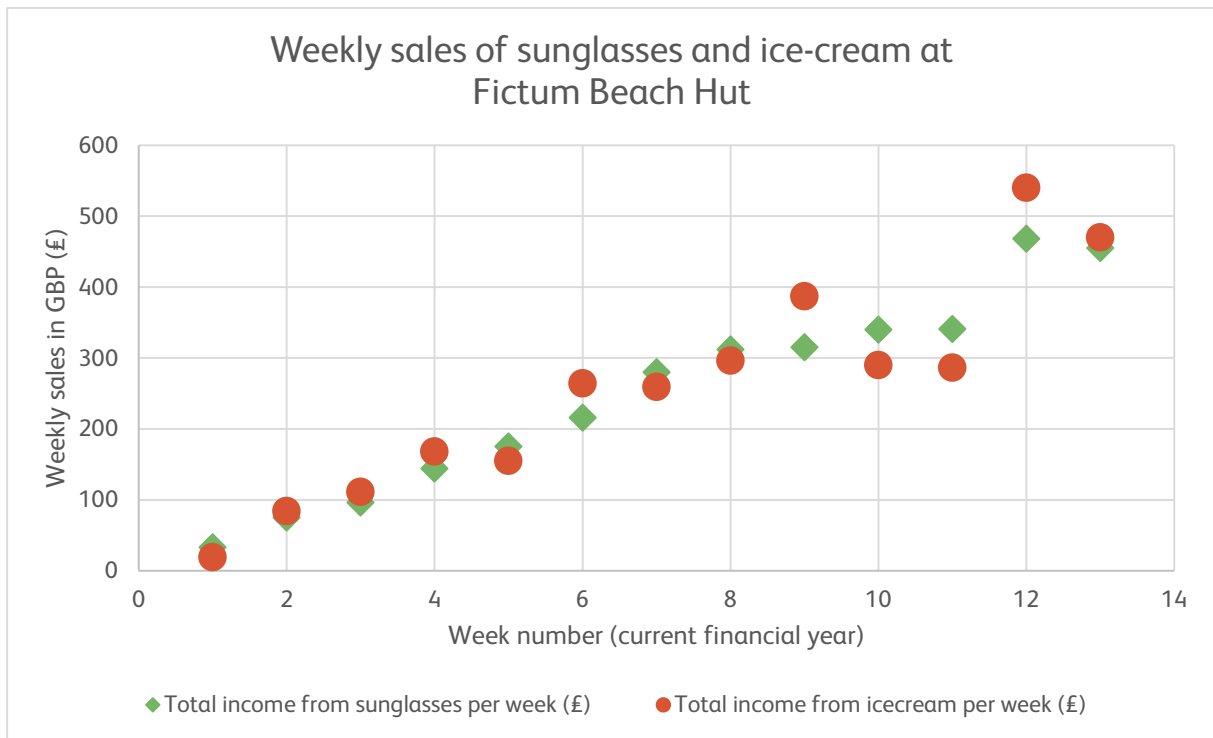
³ See U.S. National Library of Medicine, ‘[The Reports of the Surgeon General: The 1964 Report on Smoking and Health](#)’ (*Profiles in Science*, undated) <<https://profiles.nlm.nih.gov/ps/retrieve/Narrative/NN/p-nid/60>> accessed 5 October 2017



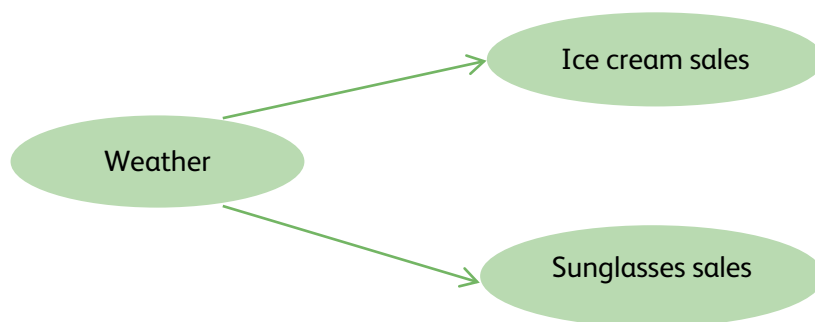
You have to be careful when considering correlation. Even if two variables seem to have a strong relationship, one variable may not in fact be the cause of another. This is what is called a **'spurious' correlation**.

An amusing website by [Tyler Vigen](#) provides a range of examples, one of which, for example, shows that per capita US cheese consumption correlates with death from becoming tangled in bed sheets in the USA. The correlation coefficient is 0.95 which is quite high. Any such observed link might just result from two independent trends that are going in similar directions.

An observed correlation might also be caused by another or **confounding variable** that affects both. Consider this simple example of a plot of values of weekly ice cream sales against weekly values of sunglasses sales.



Even if the sale of sunglasses and sales of ice cream seem to vary together, it is obvious that one does not cause the other. Rather, the weather causes changes in the sales of both.



You cannot just dismiss correlations as not indicating a cause or direct link between two variables. It may not yet have been proven. A well-conducted statistical experiment can you give very convincing evidence of causation. Subsequent studies may provide further evidence for the hypothesis. In the case of smoking, for example, the links between smoking and lung cancer have been supported by subsequent experimental scientific studies as well as observational studies.

You may need to probe further to see if there is evidence to explain the correlation – is it accidental, or an as yet unproven but strongly suggestive relationship, or is it just caused by something else?

Epidemiology and tort litigation

Epidemiology is best understood as a public health science that looks at the causes and determinants of diseases in human populations. It is primarily used to help develop health strategies for general populations. A recent example of its application in a health policy context are the bans on smoking in public spaces across the UK owing to the various risks of harm associated with second-hand smoke. There is also a growing trend of using epidemiologic research in tort litigation where the actionable damage is a complex disease (i.e. a disease about which there are significant gaps in medical knowledge). The legal claims involved tend to fall into three categories: (1) toxic torts (where the disease allegedly results from exposure to a toxic substance); (2) medical negligence claims, e.g. where a doctor is sued for failing to diagnose the relevant disease in a timely fashion; and (3) pharmaceutical products liability under the Consumer Protection Act 1987. In the vast majority of these claims, the epidemiological evidence is used at the factual causation stage of the legal enquiry (although note that in product liability claims, it is also used to determine whether the product is defective – see *XYZ v Schering Health Care* [2002] EWHC 1420 (QB)).

The science of epidemiology is expressly concerned with probabilities. For this reason, it will be less often of use in criminal cases given the high standard of proof required, though relevant epidemiology might still properly provide support for the prosecution in, say, a medical manslaughter case if there is other evidence of guilt.⁴ In civil law, the balance of probabilities standard allows for more uncertainty than the beyond reasonable doubt standard of criminal law – the evidence presented need only persuade the court that it is more likely than not that the relevant legal issue (for present purposes, it is factual causation) has been established.

Using epidemiology to decide questions of causation is controversial.⁵ The court in a toxic tort case is applying a statistical model derived from the epidemiology to the facts of a single case. In doing so the court is using epidemiological data for a very different purpose from that of the researchers. See further in [Section 4: Current and future issues](#).

Epidemiologists employ a variety of bespoke methodologies to collect and interpret data and the calculation of incidence and prevalence rates count among the most basic of these. Methods include ecological studies (comparing groups of people separated by time, location or characteristic), longitudinal or cohort studies (tracking groups of people over time), case-control

⁴ And it might be deployed in such a case by the defendant on whom no burden of proof lies.

⁵ Reservations about its use in tort litigation were recorded by members of the Supreme Court in *Sienkiewicz v Greif (UK) Ltd* [2011] 2 AC 229 though arguably some of the concerns are based on a misunderstanding of the methodologies used by epidemiologists to derive conclusions from datasets: see Claire McIvor, 'Debunking some judicial myths about epidemiology and its relevance to UK tort law' [2013] Medical Law Review 553

studies (examining the life histories of people with cases of disease), cross-sectional studies, performing experiments including randomised controlled trials, and analysing data from screening and outbreaks of diseases (see Coggon *et al.* for a description). The type of study can affect its accuracy and reliability. A longitudinal study can be prospective whilst a case-control study is retrospective. Retrospective studies are more prone to confounding and bias.

If a statistically significant relationship is found between an agent and a health outcome, one of the methods subsequently used to determine whether that relationship is indicative of a biologically causal relationship is the Bradford Hill guidelines.

Bradford Hill guidelines

Sir Austin Bradford Hill, in his President's Address in 1965 to the Section of Occupational Medicine of the Royal Society of Medicine, gave a list of aspects of association which could be useful in assessing whether an association between two variables should be interpreted as a causal relationship. These have become known in epidemiology as the Bradford Hill guidelines and are used in debates such as whether mobile telephone use can cause brain tumours.

The aspects of association, or guidelines, are:

- **Strength:** the strength of the association, for example, the death rate from lung cancer in smokers is ten times the rate in non-smokers, and heavy smokers have a lung cancer death rate twenty times that of non-smokers.
- **Consistency:** if an association is repeatedly observed by different people, in different places, circumstances and times, it is more reasonable to conclude that the association is not due to error, or imprecise definition, or a false positive statistical result.
- **Specificity:** consideration should be given to whether particular diseases only occur among workers in particular occupations, or if there are particular causes of death. This is a supporting feature in some cases, but in other cases one agent might give rise to a range of reasons for death.
- **Temporality:** this requires causal factors to be present before the disease.
- **Dose-response curve, or biological gradient:** if the frequency of a disease increases as consumption or exposure to a factor increases, this supports a causal association. Increasing levels of smoking, or of exposure to silicon dust associated with increased frequency of lung disease supports the hypothesis that smoking or silicon dust causes lung disease.

- **Plausibility:** if the causation suspected is biologically plausible. However, the association observed may be one new to science or medicine, and still be valid.
- **Coherence:** a cause and effect interpretation should not seriously conflict with generally known facts of the development and biology of the disease.
- **Experiment:** sometimes evidence from laboratory or field experiments might be available.
- **Analogy:** if there is strong evidence of a causal relationship between a factor and an effect, it *could* be fair to accept “slighter but similar evidence” between a similar factor and a similar effect.

Hill himself was keen to stress that “*none of my nine viewpoints can bring indisputable evidence for or against the cause-and-effect hypothesis and none can be required as a sine qua non*” (Hill 1965). Conversely, if none of the guidelines is met, one cannot conclude that there is not a causal association. The conclusion is that there might be a direct causal explanation, or an indirect explanation, or even that the association arose from some aspects of data collection or analysis. Epidemiologists may also consider competing explanations, such as an unmeasured confounding factor, an alternative factor which has an association of similar strength to the putative causal factor. For example, high alcohol consumption is associated with lung cancer. As people who drink often also smoke, before no smoking areas were common, pubs were often full of smoke. One should consider if the effects of drinking and smoking can be distinguished.

References and further information

D Coggon, Geoffrey Rose, DJP Barker, *Epidemiology for the uninitiated*, (Fourth edition, BMJ Books). Available online at <www.bmj.com/about-bmj/resources-readers/publications/epidemiology-uninitiated>

N Duan, RL Kravitz, CH Schmid. ‘[Single-patient \(n-of-1\) trials: a pragmatic clinical decision methodology for patient-centered comparative effectiveness research](#)’ (2013) 66(8 0) Supplement Journal of Clinical Epidemiology 21 Author manuscript available at <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3972259/>

Austin Bradford Hill, ‘The Environment and Disease: Association or Causation?’ [1965] Proceedings of the Royal Society of Medicine 295

Claire McIvor, ‘Debunking some judicial myths about epidemiology and its relevance to UK tort law’ [2013] Medical Law Review 553

Rod Pierce, ‘[Correlation](#)’ (*Math Is Fun*, 27 Jan 2017) <www.mathsisfun.com/data/correlation.html> accessed 6 October 2017

Alex Reinhart, ‘[Huff and puff](#)’ (2014) 11(4) Significance 28

Sienkiewicz v Greif (UK) Ltd [2011] UKSC 10; [2011] 2 AC 229

Tyler Vigen, ‘[Spurious Correlations](#)’ (*Spurious Correlations*, undated) <<http://tylervigen.com/spurious-correlations>> accessed 12 October 2017

For more on smoking, correlation and causation, see the *Financial Times* article [Cigarettes, damn cigarettes and statistics](#) by Tim Harford (10 April 2015), which shows how evidence of causality in relation to the impacts of cigarette smoke on health was built over time from correlation, through the timing of effects and relative dose and impact levels.

1.6 False positives and false negatives

Tests are often used in forensic science and medicine to determine whether a person has a particular characteristic, such as a blood type or an illness. Forensic tests may be carried out, for example, on products such as food, or other consumer goods, or construction materials. Like any measurement process, they are subject to error which can arise from many sources – bias, inappropriate calibration, human error mixing samples, and random variation. It is important to know what this error might be, since a test, say for evidence, might incorrectly identify a match when there isn't one, or alternatively not identify a match when there is one. In relation to medicine, a test for cancer might not show it is there when it is or, alternatively, might identify cancer when it is not there.

In 2003, a student was arrested in Philadelphia airport. Her luggage contained three condoms filled with a white powder which a field test indicated was cocaine. The student argued that these were flour and she used them as stress relief toys for exams. She spent the next three weeks in jail on drug charges until an attorney encouraged the state prosecutor to arrange laboratory testing. The tests confirmed that the powder was flour. Ultimately, she was awarded \$180,000 compensation by Philadelphia. She had been arrested and held in jail on a 'false positive' result.

With scientific or medical tests, a **false positive** result is one in which the test gives a positive result indicating the *presence* of the substance or disease for which the test was conducted when, in reality, that substance or disease *is not* present. A **false negative** result is one in which the test gives a negative result indicating the *absence* of the substance or disease for which the test was conducted when in fact the substance or disease *is* present.

The rates of false positives and false negatives (a measure of the reliability of the testing procedure) can be used to assign probabilities of observing false positives and false negatives as if the tests were applied across a population. Even very reliable tests (those with a low false positive rate) could lead to declared matches being false for many test results. This example illustrates this.

Imagine 1000 people. 10 have a disease (a prevalence of 1 %). 990 don't have the disease. All 1000 people are tested for that disease at the same time. The test used is known to have a false positive rate of 10 % and a false negative rate of 20 %. A false negative rate of 20 % means that of those 10 with the disease, only 8 are correctly identified as having the disease. The false positive rate of 10 % means that this test will identify wrongly 10 % of those 990 who do not have the disease as having the disease. At the end of the test across all 1000 people, we are therefore left with 107 people testing positive (8 + 99), but only 8 need treatment (that's about 7 % of

107); and 2 out of 893 need treatment but are not going to get it. This table summarises these figures:

	Have illness	Don't have illness	Totals
Positive test result	8	99	107
Negative test result	2	891	893
Totals	10	990	1000

Doctors have had their understanding of false positive and false negative rates tested by being presented with the above information on the false positive rate and false negative rate, and being asked: “If 1 % of the population have the illness and you get a positive result, what are the chances you have the disease?” In a scenario such as the one above, the doctors often returned an answer of about 80 %. As you can see above, that is wrong; it is about 7 % (i.e. 8 divided by 107).

The doctors answering 80 % were focusing only on whether having the illness would give you a positive test result, or not. The fact that 80 % of people with the illness get a positive test result is not the same as saying that 80 % of people with positive results have the illness.

1.7 Prosecutor’s fallacy and defence fallacy

These two misunderstandings of probability have received much attention in legal circles. They are important to understand since they can lead to incorrect interpretations of evidence or probability.

This is an area where it is worth spending a little time getting the issues clear in your head since the results can be counter-intuitive. You need to be aware how easily many people, including experts, can misinterpret what is being said.

Both the prosecutor’s fallacy and the defence fallacy incorrectly value probabilistic evidence, in part because alternative explanations are ignored. We discuss them separately here due to the burden of proof. The prosecutor’s aim is to prove guilt beyond reasonable doubt. The aim of the defence is to raise reasonable doubt. Note though, that the names are misleading – either party in court, or an expert witness, may commit either fallacy.

The prosecutor's fallacy

The prosecutor's fallacy confuses the probability of finding the evidence on an innocent person with the probability that a person on whom the evidence is found is innocent.⁶ This can equate to assuming that the probability of the scientific evidence (the match) given the defendant is innocent is equal to the probability that the defendant is innocent.

An example illustrating this, attributed to John Maynard Keynes, and discussed in Balding and Donnelly (1994) is as follows. You are playing poker against the Archbishop of Canterbury. On the first hand, he deals himself a straight flush.⁷ There are two questions you might ask at this point:

1. What is the probability of the Archbishop dealing himself a straight flush (the evidence) if he were playing honestly (i.e. innocent)?
2. What is the probability that the Archbishop is playing honestly (i.e. innocent), given that he has dealt himself a straight flush (the evidence)?

2,598,960 five-card hands can be dealt from a pack of 52 cards, of which 36 hands would be a straight flush. This means the answer to the first question is 36 in 2,598,960, or about 1 in 72,000. For the second question, most people would assess the probability of the Archbishop of Canterbury's honesty, even with the evidence that has dealt himself a straight flush, as being close to 1. The prosecutor's fallacy would be to argue that the probability of the Archbishop playing honestly given that he has dealt himself a straight flush, is 36 in 2,598,960.

To quote Balding and Donnelly (1994):

...it is possible for the two questions to have very different answers. In particular a very small answer to the first question does not necessarily imply a small answer to the second question.

The defendant is on trial for having fraudulently won the lottery by invisibly hacking into the lottery computer in such a way that it produced his number as the winning ticket. He denies any such conduct. The prosecution states that as the probability of winning is one in 20 million, then the probability of innocence is one in 20 million. This is a clear example of the prosecutor's fallacy.

Guidance for expert witnesses, particularly in forensics, encourages experts not to make such claims in written statements. However, particularly in cross-examination, or in questions to experts, in an attempt to simplify or summarise the evidence given, the fallacy may arise.

⁶ This is technically known as **transposing the conditional**, i.e. swapping over what is assumed or given.

⁷ Five cards of the same suit in sequence. An ace-high straight flush is known as a Royal flush and is excluded from the straight flush figure above.

R v Doheny

An example of the prosecutor's fallacy was elicited from an expert witness in *R v Doheny*. Doheny was accused of raping a woman in her home. She had not seen her attacker's face. The only physical evidence was semen stains found on her clothing. An expert presented evidence of an analysis of the DNA of the semen stain and blood group information obtained from the semen stain, and DNA and blood group information from Mr Doheny. In court, he was asked about the match probability of the DNA evidence combined with blood group evidence. His response was: *"taking them all into account, I calculated the chance of finding all of those bands and the conventional blood groups to be about 1 in 40 million"*. This probability is the probability of finding that set of DNA and blood type in the general population.

He is then asked by the prosecution: *"The likelihood of it being anybody other than Alan Doheny?"* He answered, *"Is about 1 in 40 million"*.

Here, the expert is being asked, what is the probability that Doheny is not the source of the semen, given the matching DNA profiles and blood type? He implicitly reverses the conditional when he answers: *"Is about 1 in 40 million."* The question posed by the prosecutor was a leading one, and elicited the prosecutor's fallacy. The appeal was allowed. The Court of Appeal in *R v Doheny and Adams* stated that an expert *"should not be asked his opinion on the likelihood that it was the defendant who left the crime stain, nor when giving evidence should he use terminology which may lead the jury to believe that he is expressing such an opinion."*

Instances of the prosecutor's fallacy can be subtle. Consider these two sentences, where 'match' refers to an agreement between a DNA profile obtained from a defendant, and a DNA profile obtained from the evidence:

- A. "The probability of a match if the semen came from another person is one in a billion."
- B. "The probability that the semen came from another person is one in a billion."

This second sentence is an example of the prosecutor's fallacy.

The first sentence is interpreted by statisticians as a statement about the evidence, specifically the probability of the evidence of a match *given* the proposition that the semen came from some person other than the defendant. It answers the question, *"Assuming the defendant is innocent, what would be the probability of finding this match?"*

The second is interpreted by statisticians as a statement about the defendant, specifically the probability of the proposition that the semen came from some person other than the defendant (the probability that the defendant is innocent), *given* the evidence of a match. It answers the

question, “Assuming the match has been made with the defendant, what is the probability of his innocence?”

The prosecutor’s fallacy is not just an issue to look for in criminal cases; it could occur in any area of law where evidence may be given an approximate probability by an expert.

The defence fallacy

If a prosecution lawyer argued that on the basis of the probability value for an evidence match, the probability of this person being innocent is so low that there is therefore a huge implication of guilt, that would be to commit the prosecutor’s fallacy.

If the defence uses the probability value for an evidence match (e.g. one in a million people) to argue that for a large enough population (for example, of the UK), their client is only one of many people that could be guilty, *and is thus innocent due to reasonable doubt*, then they are committing the defence fallacy (sometimes referred to as the defense attorney’s fallacy).

Imagine a case in which a partial DNA profile has been obtained from a crime sample. The partial profile matches the equivalent parts of the defendant’s profile. The random match probability is given as 1 in 10,000. Given the circumstances of the crime, e.g. geography, eye witness description of the offender etc., both defence and prosecution agree that the size of the population of potential sources for the crime sample is 100,000, the defence could argue that the defendant is only 1 of probably 10 matching sources. The probability therefore that he is the source is only 10%. The problem with this inference is that it assumes equal prior probability of being the source for all 100,000 of the potential sources. Almost certainly, most of the 100,000 potential sources will have different prior probabilities because of their own particular circumstances, such as motive, age or alibi.

A worked example of the prosecutor’s fallacy and defence fallacy

The following example should make these two misunderstandings clearer. There are 10,000 people in a population, and one of these is guilty of a crime. A blood stain left by the criminal has specific antigens found in 1 in 1000 people.

	Match	No match	Total
Guilty	1	0	1
Innocent	9	9,990	9,999
	10	9,990	10,000

The table above of the different outcomes shows the probability that an innocent person chosen at random has the same specific blood antigens is $9/9999 = 0.0009$ – the number of innocent matches divided by the total number of innocent people. (Of course, if the antigens did not match, the person would be excluded from suspicion.) This is not the same as the probability of being innocent given that you match the evidence. (It is also very similar to the probability of getting a match in the general population which is $10/10,000$ – a number which could have been determined through scientific studies of a population where no-one is guilty. See [Random match probability](#) in [1.8 Expert opinion evidence](#).)

The probability of being innocent given that you match the evidence is $9/10 = 0.9$ or 90%. It is the number of innocent matches divided by the total number of matches in a population. This probability might therefore be used to imply a high probability of innocence. The defence could use it to try to show that it was unlikely that this person could be guilty. However, this argument is the defence fallacy. Note that the probability statement is true. The fallacy lies in attempting to use this probability to over-ride other evidence in the case.

References and further reading

Craig Adam, *Forensic Evidence in Court: Evaluation and Scientific Opinion* (Wiley-Blackwell, 2016) Section 10.5.5 discusses the Adams case.

Colin Aitken, Paul Roberts, Graham Jackson, [Fundamentals of probability and statistical evidence in criminal proceedings](#) (Royal Statistical Society 2010) 108

DJ Balding, P Donnelly, 'The Prosecutor's Fallacy and DNA evidence' [1994] *Criminal Law Review* 711

R. v Doheny and Adams [1997] 1 Cr App R 369, 375, 377-8

A visual explanation of the Prosecutor's Fallacy and Defence Fallacy are provided on the [Probability: Fallacies, Myths and Puzzles](#) website: at Norman Fenton, 'Did the Prosecutor get it right?' (*Making Sense of Probability: Fallacies, Myths and Puzzles*, undated) <www.agenarisk.com/resources/probability_puzzles/prosecutor.shtml> accessed 6 October 2017

Mike Redmayne, 'Appeals to reason' (2002) 65 *Modern Law Review* 19

1.8 Expert opinion evidence

Most of the analysis of the relationship between probabilistic reasoning and expert opinion evidence has been done with respect to DNA evidence. However, expert opinion evidence can include fingerprints, ballistics, or trace evidence such as fibres or firearm residues, as well as epidemiological data. Different expert opinion evidence needs to be treated on its own terms, since the underlying science and tests or available comparative databases are not comparable.

Nevertheless, the fundamental issues of probability logic and inference are common to all such evidence.

Experts should think and act logically, providing opinion that is justifiable and robust. It is useful for lawyers to understand that the opinions offered by forensic scientists can be separated broadly into two basic types, depending on the type of question that is being asked

When questions are related to the circumstances of a crime or the type of offender who might have committed it, forensic scientists will be forming and providing ‘**investigative**’ opinions. In this mode, the forensic scientist is seeking explanations by generating hypotheses from preliminary observations (at, for example, a scene of crime) and then refining these after obtaining further observations. When operating in this way, forensic scientists follow the classic scientific method (see [2.3 The scientific method](#)).

In contrast, in those situations where there is a pair (or set) of competing propositions, usually relating to the defendant, forensic scientists will be operating in ‘**evaluative**’ mode. The scientist will offer the court or tribunal an assessment of the probabilities of the observations given the truth of the competing propositions. This approach may lead to an assessment of a likelihood ratio, as outlined in further detail in [Section 4: Current and future issues](#). In this mode, the forensic scientist is still being ‘scientific’, in the broadest sense of being logical and coherent, but it is a different process from that which is conventionally known as the scientific method.

Because investigative and evaluative opinions are formed in different ways, there are different challenges that lawyers can pursue to test those opinions. Guidance on these matters can be found in the RSS guide, *Case assessment and interpretation of expert evidence* (Jackson *et al.* 2015).

In *R v Dallagher*, the suspect’s conviction for murder was based almost entirely on evidence which compared an ear-print from the defendant with one found at the scene of the crime. The Law Commission noted: “*at the time of D’s trial there was an insufficient body of research data to support the hypothesis (or assumption) that every human ear leaves a unique print and that the identity of an offender could confidently be determined solely on the basis of an ear-print comparison.*” The prosecution also relied heavily on “*subjective factors ... rather than on objectively verifiable measuring techniques*”. (Both quotes from paragraph 8.11 of its Report (below)).

The Law Commission recommended how this kind of situation should be dealt with (paragraph 8.12) in its 2011 advice in *Expert Evidence in Criminal Proceedings in England and Wales*:

Under our proposed test, the prosecution would have had to prove that the witness claiming expertise was skilled in the comparison of ear-prints and therefore qualified to provide expert evidence in a criminal trial. If the defence had then made submissions on the poor data and doubtful hypothesis underpinning the expert's proffered opinion evidence, or the judge had raised the matter independently, there would have been an enquiry into the reliability of the opinion. The judge may have been able to conclude without a hearing that the expert's opinion (that D could be identified with absolute certainty from ear-prints alone) was insufficiently reliable to be admitted. Alternatively, there would have been a pre-trial hearing on the issue, which no doubt would have led to the same conclusion. The expert would not have been permitted to give an opinion that he was "absolutely convinced" that D had left the latent print at the scene of the murder. He might, however, have been able to give a weaker opinion on similarities between the latent print and D's print (assuming the jury had required the assistance of an expert in this respect).

This approach is now found in the Part 19 of the Criminal Procedure Rules and Criminal Practice Direction 2015 V Evidence 19A, 19B and 19C.

DNA

DNA may be used as evidence to support or discredit the contention that someone was present at the scene of a crime.

The RSS guide, *Assessing the probative value of DNA evidence* (Puch-Solis et al. 2012) includes a short guide to DNA and testing procedures.

In order for this evidence to be admissible it needs to be shown that the methods used are scientifically credible and valid; that the quality control in the laboratory is rigorously maintained; that the sample itself is large enough and not contaminated or degraded; and that inappropriate inferences are not being drawn (for example, on the basis of calculations from unrepresentative databases). If a 'match' is found between DNA at a crime scene and DNA from an individual, how significant is this? There may be laboratory error or there may be other people who share that DNA.

Random match probability

The **random match probability** (RMP) is the probability of finding an evidence match within a particular population. This probability is often based on sampling, so you will also need to investigate its robustness (for example, sample size, relevant population, or representativeness).

The RSS (Aitken *et al.*, 2010) has pointed out that the RMP must not be: “*confused with the probability of obtaining another match somewhere in the population. The random match probability is the probability of obtaining a match ‘in one go’, not the probability that at least one other member of the population of interest will produce a match. The probability a particular person identified in advance will win a lottery is different from the probability the lottery will be won (by someone).*”

We have already seen where this random match probability can be wrongly equated with the probability of guilt or innocence (see [1.7 Prosecutor’s fallacy and defence fallacy](#)).

In the trial of OJ Simpson, an expert witness gave a very small probability of 1 in 57,000,000,000 in relation to a bloodstain found on a gate that matched Mr Simpson’s blood profile. This figure was used in commentary on the widely-reported case to presume guilt. However, as Jonathan Koehler (1997) pointed out in his paper on the use of such probabilities, even if the figure was reliable and valid, it provides only the random match probability. “*These tiny frequencies do not themselves tell us (a) the probability that a matching suspect committed the crimes, (b) the probability that someone other than the matching suspect committed the crime, or even (c) the probability that the someone other than the matching suspect is the source of the observed characteristics.*”

Points (a.), (b.) and (c.) in the above example are not the same. Point (c) in the example is what is called the **source probability error**. This is similar to the prosecutor’s fallacy. Remember that the prosecutor’s fallacy transposes the probability of the evidence *given* the proposition of innocence, to probability of innocence, *given* the evidence. In the source probability error, the fallacy occurs if you equate the probability of finding a match between evidence and a control sample where there is no common source, with the probability of two samples not having a common source, where a match has been found.

In the example above, the only thing that can be said is that *if* the blood did not come from the accused, and the 1 in 57,000,000,000 figure has been calculated correctly, then there is a 1 in 57,000,000,000 chance of the bloodstain sample matching the blood profile of Mr Simpson. It is only the random match probability. The error is confusing the probability of a match when the suspect is not the source, with the probability that the suspect is not the source given the match. This latter probability is the wrong conditional probability for an expert to consider. Any appraisal of the probability that the suspect is not the source given a match (the fact in issue) would require consideration of the probability of the suspect not being the source in the first place, i.e. before it is known there is a match. Statisticians may state this as the prior probability that the suspect is not the source. The probability once the evidence of a match is considered, a probability known by

statisticians as the posterior probability, would require consideration of both the prior probability of the suspect not being the source and the match probability.

Sampling

Statements used by expert witnesses in relation to evidence are often based on reference datasets. In the RSS report *Fundamentals of probability and statistical evidence in criminal proceedings* (Aitken *et al.* 2010), the authors say that it is important to be able to identify and evaluate the assumptions underneath statements such as:

“The glass submitted for analysis is seen in approximately 7% of reference glass exhibits examined in this laboratory over the last 5 years.”

“Footwear with the pattern and size of the sole of the defendant’s shoe occurred in approximately 2% of burglaries.”

The appropriateness of the dataset used for reference should be questioned. How were the footwear marks or glass exhibits chosen? Are they from a population that is relevant to the proposition in question? Is there any bias in the way items were selected and submitted from criminal investigations or did the items simply come from a survey of everyday objects in people’s homes? Datasets for the latter type of items are sometimes called **convenience samples** because they are quick, easy and relatively cheap to collect but they may not provide relevant, usable data. What population might provide a better reference dataset?

A better reference dataset would be a random sample from what is known as the ‘relevant population’. Champod *et al.* (2004) showed that the definition of the ‘relevant population’ is determined by the specification of the proposition and its alternative. Consider a case of robbery in which a car is used to convey the offenders away from the scene of the crime in the north of England. The car crashes into bollards very close to the scene. An eye-witness sees the driver get out of the vehicle and run away. A footwear mark is found on the brake pedal of the car. Assume a suspect has been caught and a match found with his footwear. What population should be sampled to provide data that would be relevant to the consideration of whether the suspect’s shoe left the mark, given that we have a match? The scientist requires data that will inform him or her the probability of obtaining that match:

- a. given the truth of the proposition that the shoe made the mark
- b. given the truth of the alternative proposition that it was some other shoe

This implies two sets of data are needed, as follows:

First, to help assign a probability of a match given the truth of the first proposition (i.e. the shoe made the mark), the scientist needs data on the reproducibility of marks made by the suspect's shoe. An expert in footwear mark examination may have data on reproducibility in general but could also conduct experiments to collect data on the reproducibility of marks with this shoe.

Secondly, to help assign probabilities of a match given the truth of the alternative proposition (i.e. it was some other shoe), it would seem that some form of data related to the frequency of occurrence of the matching footwear mark would be appropriate. Sales figures might be useful in this respect. For example, the examiner may obtain information such as "*Between April 2015 and March 2017, 100,000 pairs of shoes of the same sole pattern and size as the defendant's shoes were sold in 100 outlets across the UK.*" This statement could raise questions such as why take the UK as a whole, why those dates, how many outlets are there in the north of England?

In addition, there is a problem with presenting bald, absolute sales figures. While such sales figures do offer some useful information, they are unlikely by themselves to provide estimates of relative frequencies in a relevant population.

So, what would be an appropriate relevant population from which we could compile a relative frequency and, from that, a probability of a match? In our case example, we could use surveys of footwear that have been submitted in the course of casework within forensic science laboratories. This may be a relevant population, particularly if it is from people who had been suspected of committing a crime. But, we are dealing in our case with someone who is suspected of a specific crime that involved driving the getaway vehicle in the course of a robbery. People suspected, say, of a paedophile crime may not fall within the relevant population for this type of robbery and their footwear should not therefore be taken into account.

There is a further twist - the eye witness describes the driver as "*a young woman, short in height*". The relevant population from which to assess the relative frequency of occurrence of this particular type of footwear mark would then seem to comprise a population of short women who could be suspected of this type of crime. We are not considering a population of men; we are not considering people of any age; not considering formal, fashion shoes. Whether we actually have a dataset of footwear from a relevant population of short women who could be suspected of committing the crime is a moot point but, nevertheless, the principles of evidence interpretation (such as that recommended by the European Network of Forensic Science Institutes, ENFSI (2015)) have clearly defined the relevant populations.

References and further reading

Colin Aitken, Paul Roberts, Graham Jackson, [*Fundamentals of probability and statistical evidence in criminal proceedings*](#) (Royal Statistical Society 2010) 86

C Champod, IW Evett and G Jackson, 'Establishing the most appropriate databases for addressing source level propositions' [2004] *Science and Justice* 153

ENFSI, '[ENFSI guideline for evaluative reporting in forensic science](http://enfsi.eu/news/enfsi-guideline-evaluative-reporting-forensic-science/)' (ENFSI 2015) <<http://enfsi.eu/news/enfsi-guideline-evaluative-reporting-forensic-science/>> accessed 6 October 2017

Graham Jackson, Colin Aitken and Paul Roberts, [*Case assessment and interpretation of expert evidence*](#) (Royal Statistical Society 2015)

Judicial College [*The Crown Court Compendium: Part I: Jury and Trial Management and Summing Up*](#), (2016) paras 15-34

Jonathan J. Koehler, '[One in millions, billions, and trillions: Lessons from People v Collins \(1968\) for People v Simpson \(1995\)](#)' (1997) 47 *Journal of Legal Education* 2

Law Commission, [*Expert Evidence in Criminal Proceedings in England and Wales*](#) (Law Com No. 325, 2011)

Roberto Puch-Solis, Paul Roberts, Susan Pope, Colin Aitken, [*Assessing the probative value of DNA evidence*](#) (Royal Statistical Society 2012) Includes a short guide to DNA and testing procedures.

SECTION 2: KEY CONCEPTS IN STATISTICS AND PROBABILITY

2.1 Basic probability

What is probability?

We might not know with certainty what is going to happen next, but we can talk about the **probability**⁸ of something happening.

A simple example is that of throwing a standard die with 6 possible outcomes – 1, 2, 3, 4, 5, or 6. The probability, for example, of throwing a six can be thought of as the number of ways that that outcome might happen, divided by the total number of different outcomes (i.e. 6).

$$\text{Probability (event happening – throwing a 6)} = \frac{\text{Number of ways it could happen} = 1}{\text{Total number of outcomes} = 6}$$

The probability of any outcome can therefore be expressed as a fraction (1/6), or as a decimal between 0 and 1 (in this case, 0.17). A probability of zero ‘0’ means that an event will not happen at all, and ‘1’ means there is absolute certainty the event will happen. If the probability of something happening is 0.5 (or ½), and there are two possible outcomes, it means it is equally likely that one of two events might happen. We can also express probabilities as percentages, for example, 50%.

Of course, even if the probability of throwing a 6 is 1/6, this doesn’t mean that, every sixth throw of a die will be a 6. It just means that if you threw a die say 1000 times (providing the die is not biased in any way), you would expect to get a ‘6’ 1/6 x 1000, in other words, around 167 times.

If you did get a total number of sixes that was widely different, say 250, you might suspect that there was something wrong with the die. On the other hand, it could just be a coincidence.

Section 1.2 covers coincidences in more detail and gives an example of a well-known case which illustrates the psychological tendency for people to look for causal explanations and patterns.

If you toss a coin and it comes up heads three times in a row, what would you expect to be the result the next time you throw it? Heads or tails? The probability is unaffected. It is still 50/50.

⁸ Though words such as probability, likelihood, and chance may be used interchangeably in everyday speech, to statisticians and probability theorists they have specific technical meanings. See [Section 4: Current and future issues](#) for a discussion of likelihood.

What's the probability of one thing happening OR another?

In a group of 20 people, 6 people have red hair. So the probability of a person drawn at random of having red hair is $6/20$ or 30%. There is also a 50% probability of a random person wearing black shoes. What's the probability of someone having red hair **OR** wearing black shoes?

If some people can have red hair **AND** wear black shoes (say $4/20$, 20%) then these people would be counted twice if you just added the probabilities. The event of having red hair and the event of having black shoes are therefore **NOT** mutually exclusive.

Count the figures in Diagram 2. The probability that a person chosen at random has red hair **OR** black shoes is: $6/20$ (Probability red hair) + $10/20$ (Probability black shoes) – $4/20$ (those with red hair **and** black shoes) = $12/20$ or 60%. Note that you need to know the size of the overlap to calculate this correctly.

Diagrammatic representations of individuals' hair colour (circles) and shoe colour (triangles).

Diagram 1

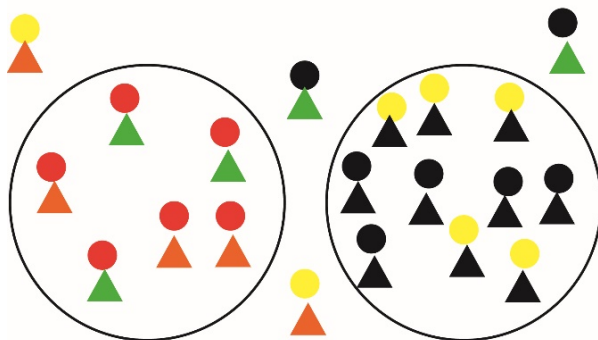
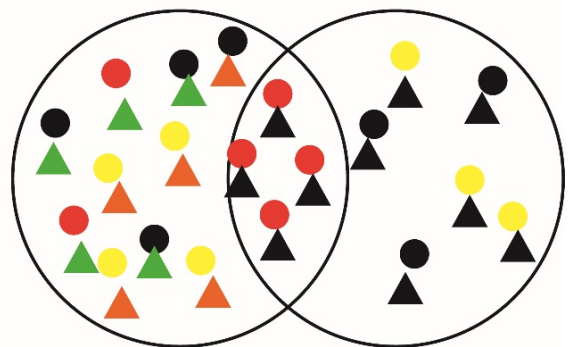


Diagram 2



What's the probability of two things happening at the same time?

What if we toss a coin and throw a die at the same time? We want to know the probability of getting a head **AND** a six. There are 12 different outcomes (see table below). Tossing a coin and throwing a die are called **independent** events. In other words, if you toss a coin, it does not affect the outcome from throwing the die. (Note that these two events are therefore not mutually exclusive – both can occur simultaneously.)

Because tossing a coin and throwing a die are independent, we can multiply the probabilities together. So the probability of getting a head **and** a 6 is $1/2$ multiplied by $1/6$ = $1/12$

We can show this visually in a table of the different outcomes. There is only one possibility (shaded) where both outcomes can happen together out of 12 possibilities.

Result of a coin flip	Result of a throw of a die					
	1	2	3	4	5	6
Head	(head/1)	(head/2)	(head/3)	(head/4)	(head/5)	(head/6)
Tail	(tail/1)	(tail/2)	(tail/3)	(tail/4)	(tail/5)	(tail/6)

Table showing combinations of a flip of a coin and a throw of a die. The one outcome of rolling a 6 and getting a head is shaded.

Events might not, however, be independent. They might be related or ‘dependent’ in some way. In this case, you can’t multiply their probabilities together. For example, two cot deaths may not be independent events but connected by, for example, genetics or some aspects of a similar environment (see [1.3 Independence](#) for further details of independence in a legal context).

2.2 Inferential statistics and probability

Inferences about the relative impact or potential causality of different pieces of evidence are a key part of legal reasoning.

Descriptive statistics provides information (a description) of the data that you have available, such as a sample (see [Refresher: The statistician’s toolbox](#)). Inferential statistics provide information on the population from which the sample was drawn.

A simple example of this is an estimate of a characteristic for a population, based on a sample. In a manufacturing plant, a quality control manager may weigh a sample of ten cakes from a production run, and take the mean of those ten measurements as an inference of the average weight of the whole production run. The figure arrived at may be correct or may be wrong by some degree (for example, if the ten picked all happened to be underweight against the specification). How confident can the quality control manager be about the figure? Inferential statistics assists with this by quantifying this uncertainty.

These inferences are always subject to uncertainty, as the wider population is uncertain. But the inferential statistics can provide useful information. The usefulness of the information depends on the quality of the information about the sample, and the **model** used by the statistician. The

model is the representation of the problem at hand, and includes assumptions, the relationship between the model and reality, and the mathematical processes used to derive the inferential statistics.

Inferential statistics in legal cases

Inferential statistics are widely used in legal cases across many areas of law. For example:

- If a cancer diagnosis has been missed, or if a person has suffered serious head injury in an accident, the settlement will usually take account of an estimate of the person's life expectancy, derived from sources such as the Ogden tables (which incorporate mortality tables kept by the Office for National Statistics) or studies of groups of similar patients. (See discussion in [Section 4: Current and future issues.](#))
- Samples of potential customers are also used in trademark infringement or deceptive advertising cases to determine the fraction of potential consumers who are confused as to the identity of the manufacturer. This typically occurs as a result of the design or packaging of a product, or a misleading advertisement suggesting that a scientific study has shown that the advertised product is superior to a competitor's.

The following sections explain how statistics and probability are used to support some simple inferences.

What's happening in the real world?

Statistics obtained from samples from a population can be used to make inferences about characteristics of the population. For example, a **sample mean** is an **estimate** of a population mean. An interval estimate for a population characteristic can also be derived which takes account of the random variability in the measurements. For a given level of confidence, the less variable the measurements, the narrower the interval.

Consider an attempt to estimate the mean height of a population of males. The mean height of the sample provides an estimate of the mean height of the population. Each time a sample is drawn from the population, a different group of people will be selected. This will change the sample mean. Statistical techniques can use the variation in heights amongst members of the sample to determine an **interval estimate** of the mean height of the population. This is an interval (a range of values) which has a stated probability of including the value of the parameter of the population.

One way to reflect how certain we are that a sample estimate reflects the value we want to know is by using a **confidence interval**, for a specified **confidence level**. If repeated samples are

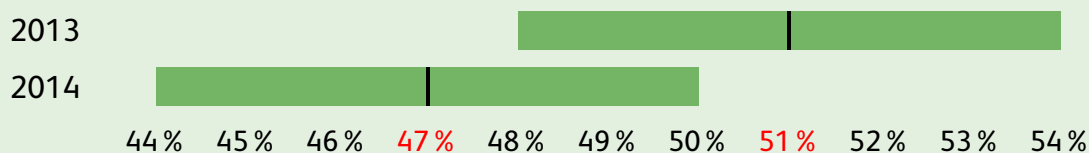
drawn from a population, then the **confidence level** is the percentage of confidence intervals that will contain the population mean. We can choose a confidence level and, from that, calculate a confidence interval of a range of values that is likely to contain the value we want to know (e.g. a range of values around the mean height of our sample, which is likely to contain the mean height of the population). It is expressed as the estimated value (for example mean height), +/- (plus or minus) a **margin of error**. The confidence level chosen and the standard deviation of the sample estimate determine the calculation of the margin of error. Typically, the wider the confidence level chosen, the larger the margin of error. Confidence intervals are usually referred to by their chosen value of confidence level, e.g. a 95 % confidence interval is one with a confidence level of 95 %. The choice of confidence level in a particular application may be specified by regulation, influenced by convention, or made by expert judgement.

An example of a confidence interval could be a survey that had a 4 % margin of error calculated for a chosen 95 % confidence level. The survey found that 28 % of respondents said they would be likely to vote for a particular party. That means that the 95 % confidence interval for the true value of what is being measured (in other words, the population percentage who voted for that party) – runs from 24 % to 32 % (in other words, 28 % +/- 4 %). Why 95 %? If 95 % confidence intervals are calculated every time this survey is performed, then 95 % of those intervals will contain the true value. For this one specific survey, we can't know whether this is one of the 95 % of intervals that contain the true value, or the 5 % that don't, but the confidence interval does give a range of plausible values that the true value might take. In opinion polls, a 95 % confidence interval is conventional. A 99 % confidence interval could be 'better' in terms of being certain that this is one of the intervals that contains the true value, but the trade-off is that the margin of error will be larger, so the results will be less precise.

Sample size matters. For example, suppose that with a sample size of 100 with a random sample you get a margin of error of about +/- 10 % then with a random sample of 1000 people that margin of error falls to about +/- 3 %.

An [opinion poll](#) leading up to the Scottish Referendum in 2014 was carried out in August 2013 and then August 2014. In 2013 the result of a question asking if Scotland should be independent was that 51 % said yes with a margin of error of +/- 3 %. The 2014 opinion poll result was 47 % +/- 3 %. So you might think that over that year there had definitely been a fall in the potential yes vote. However, look at the ranges. The actual result for 2013, with a 95 % confidence interval, could plausibly be anywhere between 48 % and 54 %, and for 2014 between 44 % and 50 %. The two ranges overlap. That gives a hint that the potential yes vote might not have fallen after all. Could it even have increased slightly? One can't tell

how plausible that is without doing some further calculations, but the uncertainty in the two original figures provides a warning about reading too much into small changes in survey results.



Confidence intervals are one measure of **sampling error**. **Non-sampling error** refers to other errors, including biases of data collection, such as leading questions in surveys, or fingerprint databases that contain only previous criminals.

Inferential statistics can also be used to assess the statistical significance of an association between two characteristics, e.g. smoking and lung cancer. For example, it can be used to assess whether or not a medical intervention has an effect, or if there are linkages or associations between data. We have explored **correlations** in [1.5 Correlation and causation](#).

What's the probability of something happening once something else has happened?

The situation of determining the probability of an event after something else has happened is called **conditional probability**. The probability of the second event happening is 'conditional' on what happens first.

If, for example, you have a bag of 100 marbles – a mix of red, green and blue – the probability of drawing a red marble is 5/100 if there are only five red marbles in the bag and all marbles are the same shape. If you draw a marble out of the bag, and do not put it back, the probability of drawing a red one the second time is changed to either 5/99 if you did not draw a red one out the first time, but 4/99 if you did.

In other words, the two events are not independent. They are **dependent**. As we saw in sections [1.3 Independence](#) and [2.1 Basic probability](#), you can't multiply probabilities together if the two events are dependent.

The misunderstanding of conditional probability is behind some well-known miscarriages of justice which we have explored in [1.7 Prosecutor's fallacy and defence fallacy](#), and [1.8 Expert opinion evidence](#). We also considered dependent and independent variables further in 1.3 Independence. We will consider them further in [2.3 The scientific method](#).

Your expert witness may write these probability relationships using specific notation.

For example, when two events A and B are mutually exclusive, then the probability of either A or B occurring is written as:

$$P(A \cup B) = P(A) + P(B)$$

(A \cup B is said as A 'union' B – a mathematical term to combine separate groups)

When two events are *not* mutually exclusive, then the probability of either A or B occurring is written as:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

(where '∩' means 'intersection')

The 'intersection' means the overlap between the group of all 'As' and the group of all 'Bs'. You must take this number away from the total, otherwise you would be double-counting.

The probability of a barrister being female and the probability of a barrister being left-handed are not mutually exclusive. Say 4 in 10 barristers are female, then the probability of picking a barrister at random and them being female is 0.4. $P(\text{female}) = 0.4$

Say 1 in 10 barristers are left-handed – that's a probability of 0.1. $P(\text{left-handed}) = 0.1$

To find the total probability of a barrister being either left-handed or female, add the probability of a barrister being female to the probability of a barrister being left-handed (0.4 + 0.1), and deduct from that the overlap of the two – the probability of a barrister being both female and left-handed (say it is 2 in 100, which is 0.02). Thus the probability of a barrister being either female or left-handed is $P(\text{female} \cup \text{left-handed}) = 0.4 + 0.1 - 0.02 = 0.48$ or 48 out of 100.

Does that mean the probability of a barrister being either male or right-handed is 0.52? No. Assuming only male and female, and left- and right-handedness, $P(\text{male}) = 0.6$, $P(\text{right-handed}) = 0.9$, $P(\text{male} \cap \text{right-handedness}) = 0.52$ $P(\text{male} \cup \text{right-handed}) = 0.6 + 0.9 - 0.52 = 0.98$.

Conditional probability is written as $P(B | A)$ which means the probability of B **given that** A has already happened (or is hypothesised to be the case). (It is important to know that $P(A | B)$ is not the same as $P(B | A)$. In other words, the probability that A has happened given that B has already happened, is (apart from very rarely) not the same as the probability that B has happened given that A has already happened.)

For example, if you look at the example in [The defence fallacy](#) in [1.7 Prosecutor's fallacy and defence fallacy](#), the probability that a person matches the evidence given that they were innocent $P(\text{match} | \text{innocent})$ is 0.0009 while the probability of being innocent given a person matched the evidence $P(\text{innocent} | \text{match})$ is 0.9.

Reference and further reading

Colin Aitken, Paul Roberts, Graham Jackson, *Fundamentals of probability and statistical evidence in criminal proceedings* (Royal Statistical Society 2010)

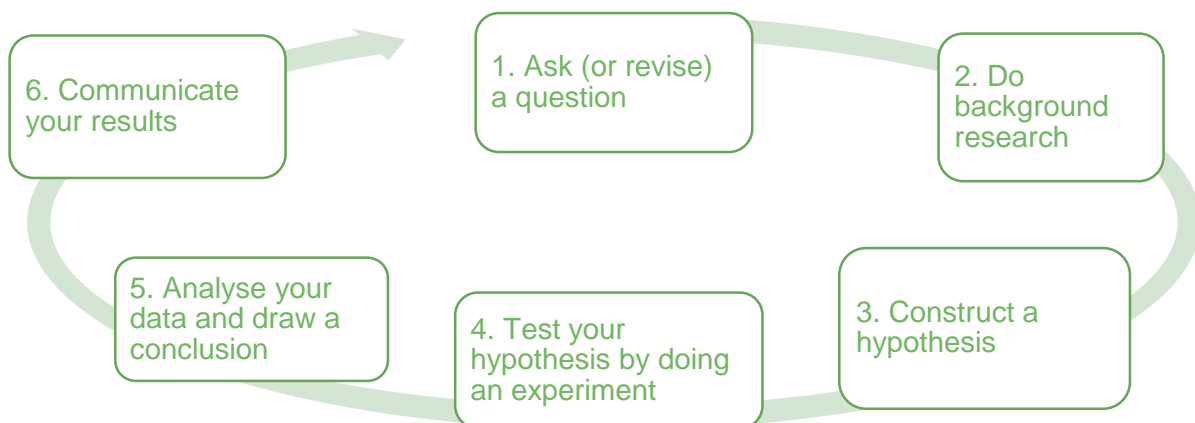
Joseph L Gastwirth, 'Statistical Reasoning in the Legal Setting' [1992] *American Statistician* 55 provides examples of inferential statistics in US and Canadian courts.

Joseph L. Gastwirth, Samuel W. Greenhouse, 'Biostatistical Concepts and Methods in the Legal Setting'. [1995] *Statistics in Medicine* 1641 describes a case concerning the pay of clerical employees in a court.

2.3 The scientific method

The scientific method is a formalisation of a way of thinking and acting in a logical, effective and efficient way. It forms the basis of a research methodology that guides the scientist in forming hypotheses, in designing experiments to collect data to test the hypotheses and in evaluating those data. Different hypotheses about, for example, causality can be explored in this way.

A simple way to consider the scientific method is as a simplified series of six steps that operates in a cycle (Ham and MaHam 2016):



One common approach used in medical research, for example, is a **randomised controlled trial**. Here, people are randomly assigned to two groups, one of which is treated in a certain way, and the other either given a different treatment or a placebo. '**Double blind**' approaches, while not always feasible, are the most rigorous since the treated and untreated group do not know which they are, and neither do the researchers.

There is also a range of non-experimental methods of conducting science such as in medical research when using correlations between data (see [1.5 Correlation and causation](#)).

Significance testing

When considering the results of a scientific investigation, the role of random chance and variability must be considered. Just as for estimates of population characteristics (see [2.2 Inferential statistics and probability](#)) there is likely to be a lot of random variation in scientific results. Is there a link between your experiment's results and the actions you took over the course of your experiment? Did your experiment 'work', or did you just have a lucky day? You can use tests of statistical inference, such as significance testing, to help address that. As we explain in the box below, statistical inference can only calculate a probability for obtaining the experimental values you obtained (or anything more extreme), if your experimental actions had no influence over the results. A small value for this probability does not imply a large probability for the hypothesis that there is an effect. A small value is a measure of support for your results, not a definitive statement about the probability of the truth of your experimental hypothesis.

In statistical tests we usually work with two hypotheses, the null and the alternative. The null hypothesis is something like the status quo; it is the assumption we would make unless there was sufficient evidence to suggest otherwise. The alternative hypothesis represents a new state of affairs that we suspect (or perhaps hope) might be true.

For example, suppose we are testing a new medical treatment to see if it performs better than the existing standard treatment. The null hypothesis would be that the new treatment is no better (or worse) than the old; the alternative would be that it performs better.

A **significance** test assesses the evidence in relation to the two competing hypotheses. A significant result is one which favours the alternative rather than the null hypothesis. A highly significant result strongly favours the alternative.

The strength of the evidence for the alternative hypothesis is often summed up in a 'P value' (also called the significance level) – and this is the point where the explanation has to become technical. If an outcome O is said to have a P value of 0.05, for example, this means that O falls within the 5% of possible outcomes that represent the strongest evidence in favour of the alternative hypothesis rather than the null. If O has a P value of 0.01 then it falls within the 1% of possible cases giving the strongest evidence for the alternative. So the smaller the P value the stronger the evidence.

Of course, an outcome may not have a small significance level (or P value) at all. Suppose the outcome is not significant at the 5% level. This is sometimes – and quite wrongly – interpreted to mean that there is strong evidence in favour of the null hypothesis. The proper interpretation is much more cautious: we simply don't have strong enough evidence against the null. The alternative may still be correct, but we don't have the data to justify that conclusion.

A measure of the strength of evidence is **power**. When we interpret the P value to say that we cannot *reject* the null hypothesis, this does not automatically mean that we should *accept* the null hypothesis. The proper response should take into account the power of the test. Power is a measure of the probability of correctly rejecting a false null hypothesis, that is, the probability of correctly detecting an effect when there is an effect there to be detected.

The power of a statistical test depends strongly on the amount of data available. It is easier to identify false null hypotheses with large samples than it is with small samples – more evidence makes better decisions possible. However, there is a danger lurking here: very large samples can result in small P values even though the effect being detected is tiny and of little or no practical importance.

The interpretation of P values is one of judgement. In many areas of research, a P value of 0.05 or below is conventionally taken to be sufficient to reject the null hypothesis. In other areas, a much more stringent figure is required. An extreme example is the work at CERN to discover the existence or not of the Higgs Boson, where they set themselves a threshold equivalent to a P value of 0.0000003. The interpretation of P values is an area of vigorous debate within the scientific and statistical communities.

There are interesting parallels here with criminal cases in a court of law. The null hypothesis in court is that I am not guilty. This is the assumption we start with; the assumption we hold to unless there is sufficient evidence otherwise. The alternative is that I am guilty, and the court accepts that conclusion only if my guilt is shown ‘beyond reasonable doubt’. But if the prosecution fails to obtain a guilty verdict this does not show that I am innocent. Perhaps I am innocent; or perhaps I am guilty but the evidence is not strong enough to show guilt beyond reasonable doubt. In the latter case, additional evidence may emerge later and I may face a re-trial.⁹

Likewise, if the evidence in favour of the new medical treatment is strong enough, we will want to adopt it. But if the evidence is weak we will stick with the standard treatment, at least until additional experimental evidence emerges to suggest that the new treatment may be better.

Reference

Bryan M Ham and Aihui MaHam, *Analytical Chemistry: A Chemist and Laboratory Technician's Toolkit* (John Wiley and Sons 2016)

⁹ In England, only in exceptional circumstances for certain offences through an application to the Court of Appeal to quash the acquittal under part 10 of the CJA 2003.

SECTION 3: INTO PRACTICE

In this section, we look at how these statistical issues may arise in practice, and how to manage them.

We start with a discussion on communicating with experts about statistical issues.

We then introduce a set of guidelines for experts, which have recently been developed by the ICCA.

We conclude by presenting four (fictional) case studies, drawn from different areas of legal practice, which throw up many of the evidential problems discussed in Sections 1 and 2 of this booklet. They are offered as suitable cases for discussion and analysis in advocacy training sessions. There is no common thread connecting them. Each poses its own set of challenges for the advocate. Can they be identified in advance, and how are these different challenges answered?

- **Case study one: criminal.** An example of a robbery involving a murder, where DNA evidence is available from the suspected murder weapon and from the crime scene. How should you scrutinise the evidence to assist with the case for the defence?
- **Case study two: civil (medical negligence).** A case where the claimant was not referred to hospital when they saw their GP. The court is to consider evidence as to whether the GP's inaction would have made any difference to the claimant's current condition. What questions should be asked of the expert witnesses in the case?
- **Case study three: family.** A case of a care order being made with respect to a child. The court will consider the current and future capability of the mother. Evidence about the mother's alcohol intake will form part of the court's decision. Is it reliable?
- **Case study four: planning.** Involves a developer wanting to build a new supermarket, against some residents' wishes. A professor of mathematics has provided evidence relating to traffic flow, and local volunteers have conducted a survey. Will they help or hinder your case?

3.1 Communicating with experts about statistical evidence

Are you speaking the same language?

Statistical language is not necessarily the same as legal language, and neither are necessarily the same as everyday language. Words that you use may have different meanings for jurors, experts and lawyers. Statisticians can use what might appear to be an everyday word as a specific

technical term. Take **significance**, for example. In everyday language, significance carries associations of importance, something with considerable meaning. In statistics, significance is a measure of whether a result is due simply to chance or random variability or not (see [2.3 The scientific method](#)) and specifically does not indicate whether the result is of importance or benefit.

In addition, the same word may be used in statistics as in another professional discipline, but with a different intended meaning. Take significance again. A clinician may use significance as shorthand for clinical significance, a measure of whether something is of practical value to a patient.

We **highlight** several terms with specific statistical meaning in this guide, many of which also have everyday meaning. Further examples are provided in McConway (2016) and Aitken & Taroni (2008).

How expert is your expert?

The Sally Clark case (see [1.3 Independence](#) for an outline) is a famous example of where an expert witness (in medicine) incorrectly handled statistics in court. Guidelines for expert witnesses (see below) require experts to comment only within the boundaries of their expertise. This assumes, of course, that they are aware of the boundaries of their expertise, even under cross-examination. The Royal Statistical Society offers Chartered Statistician (CStat) as a professional award, which may be expected for statistical expert witnesses.

Statisticians are not always certain

Statistics and probability apply scientific techniques to uncertainty. They draw inferences to describe uncertain situations, and make decisions in circumstances of uncertainty.

Statisticians use judgement and skill, as well as calculation and models. A statistical problem requires the statistician to look at what questions can be asked, what data are available or collected, what assumptions or existing conventions or prior work to consider, how to analyse the data appropriately, and how these results can be applied to the questions and the original problem. These choices should be based on current conventions and good practice, and be justifiable. The resulting calculations will only be as reliable as the model they are derived from.

As David Spiegelhalter comments, *“In general, I don't feel statistical evidence is handled well by courts. They like either incontrovertible numerical “facts”, or overall expert opinions. But statisticians deal with a delicate combination of data and judgement that often gives rise to*

“rough” numbers, and these don't seem to fit well with the legal process.”¹⁰ Statisticians are not merely number-crunchers; they use interpretation and judgement to discharge their professional duties.

The questions the statisticians can answer may not be the questions you ask

Statisticians can only work with what's available, and may want to address simpler questions than those originally posed, as that is what the available evidence allows (see Bird and Hutton (2012) for an example of this).

In order to discharge their duties to the court properly as expert witnesses, experts need to be clear about the questions or propositions that they are being asked to address and about which of those are within their remit and competence to answer. There will be questions for which the court needs answers but for which it would not be safe or logical for the expert to provide those answers. A case involving an assault by kicking provides an example. The prosecution allege the defendant kicked the victim repeatedly while the victim was lying on the floor of an alleyway in a busy city centre. The defendant says he had been drinking heavily on the night in questions and he remembers walking down the alleyway and stumbling over a body lying on the floor. The expert gave evidence to the effect that the pattern of blood-staining observed on the defendant's shoes and trousers was *far more likely* to be seen if the kicking allegation were true rather than if the stumbling alternative were correct. The expert bases that opinion on answers to questions of the type – how often would you expect to see blood-staining of this type on people who have kicked bodies compared to people who have stumbled over bodies? She can rightfully provide expert answers to these questions. She cannot answer the question – given your observations of the blood-staining, how likely is it that the defendant kicked the victim? The answer to that question requires knowledge of, and an opinion about, the other evidence in the case. The expert may have very little or no knowledge of the other evidence or, if she does have some knowledge, she may have formed a biased view of the probability that the defendant kicked the victim.

Differences of approach

As a starting point to understanding basic statistics and how to use them in legal contexts, advocates need to be able to appreciate the similarities and differences between how lawyers and statisticians think about risk, probability and causality. Correspondingly, scientists who understand the implications of the questions being asked of them will make better expert witnesses.

¹⁰ Oz Flanagan, 'Statistics and the law: a recent history of an uneasy relationship' (StatsLife, 13 February 2015) <www.statslife.org.uk/features/2819-statistics-and-the-law-a-recent-history-of-an-uneasy-relationship> accessed 6 October 2017

Both lawyers and statisticians deal with the inferences and consequential decisions which can safely be reached in situations of uncertainty.

Statisticians, and experts who use statistics, whether they are examining the probable cause of some past event, or making a prediction as to what is likely to occur in the future, will refer to general data and other similar information, and consider the inferences which can properly be drawn from that material. For example, they might look at the frequency of a genetic marker in the general population or the frequency of the occurrence of an event such as a cot death or an engine failure. This will create a context in which a set of similar facts – past or future – can be evaluated. Often this will involve the expert drawing on statistical evidence, combined with their own judgement. An expert on fire, for example, may use evidence from previous fires to draw up a list of potential causes of fire in the current case, and then use observations of the current case to assess the probable cause(s).

The decisions of the courts, by contrast, are determined by applying a burden and standard of proof. The standard of proof is described as ‘beyond reasonable doubt’ in criminal cases and the lesser ‘on the balance of probabilities’ standard in all civil litigation. Neither term has a technical scientific meaning but both are routinely applied by judges and juries in all cases involving factual disputes.

Lord Mackay in *Hotson v East Berkshire Area Health Authority* [1987] A.C. 750, discussing the use of statistics as an aid to proof of causation, referred to the Californian case of *Herskovits v Group Health Cooperative of Puget Sound* and the dissenting judgment of Justice Brachtenbach who considered the following. If you were knocked down by an unknown taxi cab in a town with two cab companies, it would (other things being equal) be more probable that you would have been knocked down by the one with the larger fleet. Mackay concurred with Brachtenbach’s assessment that that is not enough evidence for any court to find that company liable.

Lord Phillips in *Sienkiewicz v Greif (UK) Ltd* described this as “an extreme example of the fact that statistical evidence may be an inadequate basis upon which to found a finding of causation.”¹¹

The economist John Kay summed this up pithily as:

A court is concerned to establish the degree of confidence in a narrative, not to measure a probability in a model. (Kay 2013)

While the standard of proof required to discharge that burden is never less than the balance of probabilities (the civil burden of proof: “more likely than not”), statistics which do not themselves

¹¹ [2011] UKSC 10; [2011] 2 AC 229 [96]

disclose a greater than evens chance of something happening may still assist a party on whom a burden of proof lies: thus in cases where the law recognises that a claimant is entitled to damages for the “loss of a chance”¹², a statistically evidenced chance of less than 51 % may still have probative value. For example, an injured claimant whose injury has prevented him continuing his career may seek to rely on data showing that 20 % of those at his level in the organisation go on to achieve promotion to a higher grade in order to enhance his claim for loss of earnings by a similar percentage. (However, if it is not admitted, the existence of the lost 20 % chance must itself be proved by the claimant to the balance of probabilities standard.)

We can see that the approaches of the expert and the advocate differ. This is not to say that either is “wrong”; the legal approach to evidence and causality has a long history and is generally effective. As regards statistical information, the lawyer’s line of enquiry is much narrower than that of the statistician. Defensible legal interpretation of expert statistical opinion is possible. The conceptual issues of legal standards of proof should be understandable to statistical experts, and conceptual clarity will enable lawyers to properly apply the standards of proof.

Further developments

People understand and interpret statistical and probabilistic evidence in different ways. It is important that expert witnesses are credible in their use of statistics and able to present their knowledge and opinions clearly and appropriately and that advocates have a clear understanding of the way they work.

Particularly where DNA and other forensic evidence is used, there is ongoing development of how to handle and present evidence using statistical techniques (See particularly [1.8 Expert opinion evidence](#) and [Section 4: Current and future issues](#)).

References and further reading

Sheila M Bird and Jane L Hutton, ‘[Statistician expert witnesses in agreement on relative hazards](#)’. (2012) 18(2) Clinical Risk 58

Philip Dawid, ‘[Statistics on Trial](#)’ (2005) 2(1) Significance 6

A Philip Dawid, ‘The role of scientific and statistical evidence in assessing causality’ in Richard Goldberg(ed) *Perspectives on Causation* (Hart Publishing 2011) and A. Philip Dawid, David L. Faigman and Stephen E. Fienberg, ‘*Fitting Science Into Legal Contexts: Assessing Effects of Causes or Causes of Effects?*’ [2014] Sociological Methods & Research 359 both discuss differences between science and law in terms of the questions asked about causes on populations and individuals.

Gregg (FC) v. Scott: UKHL 2; [2005] 2 AC 176 [79].

[Herskovits v. Group Health Cooperative of Puget Sound](#), 99 Wn.2d 609, 664 P.2d 474, 1983 Wash.
<https://h2o.law.harvard.edu/cases/5439>

¹² Such as in *Allied Maples Group Ltd v Simmons and Simmons* [1995] 1WLR 1602 CA.

ICCA *Guidance on the preparation, admission and examination of expert evidence* (ICCA 2017) Available from www.icca.ac.uk

John Kay, '[A story can be more useful than maths](#)', (*Financial Times*, London, 11 February 2016)

Kevin McConway, '[More confusing terms in statistics](#)', (*Open University OpenLearn*, 11 February 2016) <www.open.edu/openlearn/science-maths-technology/mathematics-and-statistics/statistics/more-confusing-terms-statistics> accessed 6 October 2017

3.2 Guidelines for experts

In the discussion of these case studies, the following Guidelines (GLs) are likely to be relevant. They summarise principles which are now well-established across all areas of advocacy. Many of these guidelines originate in the judgment of Cresswell J in the Commercial Court in *The Ikarian Reefer* [1993] 2 Lloyd's Rep 68 at 81-82. Later rules of court and good practice have expanded and enlarged upon the principles stated in that case. They are reflected in the Civil Procedure Rules, Part 35 and Practice Direction 35, and the recommendations of the Law Commission (2011) in *Expert Evidence in Criminal Proceedings in England and Wales* (Law Commission No. 325), now incorporated in the Criminal Procedure Rules and Practice Directions, Part 19.

GL1 - An expert witness owes a duty to the court to give independent, objective and unbiased evidence within his or her area of expertise.

GL 2 - The expert's duty to assist the court overrides any duty owed to the party by whom the expert is instructed or paid.

GL 3 - An expert owes a duty to the court to define his or her area of expertise and inform the court of any question to which the answer would fall outside his or her area of expertise.

GL 4 - An expert must make clear which facts relied upon are within his or her own knowledge and which facts are derived from other sources.

GL 5 - Where any facts, including examinations, measurements, tests or experiments, have been provided or carried out by others, the expert must say from whom the relevant information has been obtained and the extent to which (if at all) the expert participated in the obtaining of the facts or material in question.

GL 6 - Experts should always resist any attempt by advocates to present their opinions in numerical form, and certainly not by reference to the relevant standard of proof.

GL 7 - Where there is a range of opinion on any matter the expert must summarise it and explain why he or she has reached his or her own conclusion.

GL 8 - Where there are material facts in dispute, the expert should give his or her opinion on each hypothesis and should not express a view in favour of one version or another unless, by virtue of expertise and/or experience, he or she can express and justify a view on the probabilities.

GL 9 - If at any stage in legal proceedings an expert believes that there is a reason for changing or qualifying his or her opinion, the court and the parties must be informed immediately.

GL 10 - Any of the following could provide a reason for determining that expert opinion evidence is not sufficiently reliable —

- a. the opinion is based on a hypothesis which has not been subjected to sufficient scrutiny (including, where appropriate, experimental or other testing), or which has failed to stand up to scrutiny;
- b. the opinion is based on an unjustifiable assumption;
- c. the opinion is based on flawed data;
- d. the opinion relies on an examination, technique, method or process which was not properly carried out or applied, or was not appropriate for use in the particular case;
- e. the opinion relies on an inference or conclusion which has not been properly reached.

GL 11 - In assessing the reliability of expert opinion evidence, the court will have regard to:

- a. the extent and quality of the data on which the opinion is based, and the validity of the methods by which they were obtained;
- b. if the opinion relies on an inference from any findings, whether the opinion properly explains how safe or unsafe the inference is (whether by reference to statistical significance or in other appropriate terms);
- c. if the opinion relies on the results of the use of any method (for instance, a test, measurement or survey), whether the opinion takes proper account of matters, such as the degree of precision or margin of uncertainty, affecting the accuracy or reliability of those results.

Read through each case study carefully and firstly work out what these statistics and probability issues are, and what may be problematic about each. Then work out an appropriate line of questioning to use in each case.

Further reading

ICCA *Guidance on the preparation, admission and examination of expert evidence* (ICCA 2017) Available from www.icca.ac.uk

3.3 Case study one – Criminal

Read the following scenario. You are acting for D in this case.

D is accused of murder. Three men entered the deceased's premises, which was a first floor flat, accessed from a front door giving out onto the street. He was fatally stabbed. The motive was believed to be robbery. An eye witness saw three men ring the bell on the street and being let in by the deceased. The same witness saw three men run from the premises a few minutes later. No useful descriptions were noted.

A knife, believed to be the murder weapon, was found discarded in a nearby dustbin a few hours later. The knife had been brought to the scene, i.e. it was not from the flat of the deceased. It was sent for forensic analysis to a private provider of forensic services.

The doorbell was also swabbed for DNA, and the swabs were sent for analysis.

There is a prosecution report as to DNA findings on the knife and doorbell.

These state that there is blood on the blade of the knife, which yielded a full single DNA profile. That profile matched the profile of the deceased. The probability of the match being obtained if the blood originated from someone else unrelated to the victim was said to be in the order of 1 in 1 billion.

There was a complex mixed DNA profile on the handle of the knife. The report states that there were at least 3 contributors to the mixture. It further states that the profile was run through a computer programme used by the forensic provider. It sets out two competing propositions. H1 is that the mixture is from D plus two unknown contributors (the prosecution hypothesis). H2 is that the mixture comes from 3 unknown contributors, not including D (the defence hypothesis). The report states that the DNA profiling results are 1,000,000 times more likely under H1, than under H2. This is said to provide "very strong support" for the presence of DNA from D in that mixture.

As to the doorbell, the prosecution report states that there is again a complex mixed DNA profile from at least three contributors, but that the mixture is not amenable to a statistical evaluation because of the poor quality of the mixed profile. It, however, states that the majority of the components of D's profile are found within the mixture, and that the results are such as the expert might expect to find had D contributed to that mixture. It goes on to state that the findings relating to the doorbell provide strong support to the proposition that D contributed DNA to the doorbell mixture rather than the proposition that he did not.

D is arrested two months later. There is some other circumstantial evidence of involvement, but none that would found a case to answer, absent the DNA. D denies any involvement, and says that any match must be purely coincidental.

Points for discussion

- What further disclosure should you ask for in relation to any of the findings?
- Would you want to instruct your own expert and, if so, what specific questions might you ask him/her?
- Are there any admissibility issues? If so, what are they?
- What challenge can you make to any of these conclusions in cross-examination?
- Has the opinion in relation to the doorbell findings conformed to the relevant Guidelines (14-16) in the Forensic Science Regulator's guidance on reporting DNA mixtures? In particular, has the prosecution established that the expert has sufficient relevant/audited experience to give a reliable, qualitative opinion on the strength of the evidence? Should this opinion be excluded?
- Use of pt 19 CrimPR and CPD – have these been complied with?

Further reading

Forensic Science Regulator, *Legal obligations: issue 5. FSR-I-400* (Forensic Science Regulator 2017)

<<https://www.gov.uk/government/publications/legal-obligations-issue-5>>

Forensic Science Regulator, '*DNA mixture interpretation: draft guidance*' (Forensic Science Regulator 2017).

Consultation on draft guidelines underway at time of going to press: <www.gov.uk/government/consultations/dna-mixture-interpretation-draft-guidance> accessed 6 October 2017

3.4 Case study two – Civil (medical negligence)

Read the following scenario. You are acting for a claimant (C). You are preparing to see your side's expert witness Dr X in conference. Consider the points you need to discuss with him (including any weaknesses in his evidence). Consider also what points you would want to put to the other side's expert witness (Professor Y) in cross examination.

The undisputed evidence is that, from 7 am, C then aged 45, in previous good health though with a family history of hypertension, started to experience mild symptoms which it is later established were the precursors of a stroke. He saw his GP at 9 am but the GP (D) gave reassuring advice and prescribed no treatment. Six hours later, at 3 pm, C collapsed with a massive stroke and is now severely disabled.

C sues D and it is admitted on the latter's behalf that he was negligent in not making an emergency referral to hospital. It is also admitted that, had that been done, C would have got to a specialist stroke unit within an hour, have undergone a CT scan within another hour, which would have shown the developing blood clot in a cerebral artery, and, in consequence of this finding, C would have been started on an anti-thrombotic drug by 11.30am at the latest.

But it is said on behalf of D that nevertheless C would still not have avoided the massive and irreversible deterioration at 3pm and therefore the negligence of the GP made no difference. So there is an issue of causation. Essentially the rival cases put by the neurologist/stroke physicians called as expert witnesses on each side are as follows:

C's expert Dr X says that the literature shows that early intervention within five hours of first symptoms more than halves the chances of severe stroke. D's case based on Professor Y's opinion is that causation is not proved because a different study shows that early intervention with an anti-thrombotic reduces the percentage chances of recurrent stroke by only a very small amount and no more, in C's case, than 10%.

Dr X relies on a study of 850 stroke patients between 2006 and 2008 which shows that after one month 80% were in a better condition than C is now in and only 20% were in as bad a condition or dead. He says this supports his view, based on his own extensive clinical experience, that with proper and timely treatment in hospital, C's chances of avoiding deterioration would have been more than doubled.

Professor Y says that this study does not assist in understanding how the failure to treat C in this case made a difference to the outcome because the study included no “controls” who were not given any treatment.

Professor Y relies on a retrospective observational study of 1000 patients treated both at home and in geriatric hospitals between 2008 and 2010 which shows a much greater likelihood of deterioration whatever course of treatment was prescribed.

Both experts also refer to a randomized control trial of 20,000 patients admitted with minor symptoms of stroke to general hospitals around the world carried out over ten years from 1987. Half were given an anti-thrombotic within three hours of admission and half were given a placebo.

Dr X says this shows a significant reduction in the risk of recurrent stroke and, in view of improved standards of care in modern specialist stroke units, it supports his view that C’s chances of a better outcome would have been more than doubled if he had been given anti-thrombotic treatment by 11.30am.

The report of this study published in the *British Medical Journal* shows that the risk of severe stroke within 24 hours of admission was very low even among those who were given no drug treatment (500 patients or 5%). But it improved to (300 or 3%) for those given an anti-thrombotic.

Professor Y says that this study in fact therefore shows that the treatment only benefitted two people for every 100 treated with the anti-thrombotic so that there was only a 2% reduction in risk. Because C would have been cared for in a specialist stroke unit and thus received optimum care, Professor Y was prepared to increase the reduction in risk in C’s case from 2% to 10% but not more, still nowhere near enough to support a case that C would, on the balance of probabilities, have avoided his severe stroke.

Dr X says a correct interpretation of this study in fact shows a very much higher reduction in risk than 2% – he says the relative risk reduction supports his view that C would probably have avoided deterioration if he had received an anti-thrombotic at 11.30 am.

Both experts agree that treatment for suspected stroke has much improved in the last decade.

Points for discussion

- Should you challenge the type, quality and reliability of the various studies referred to by the experts?
- How do you do that? See [2.3 The scientific method](#). What are your conclusions?

- Have the experts adequately addressed the distinction between absolute and relative risk? See [1.4 Absolute and relative risk](#).
- Is there a confusion between correlation and causation? See [1.5 Correlation and causation](#).
- Is it appropriate to make adjustments to the conclusions drawn from these studies for the claimant's age and previous good health status; if so, how can that be done? What about the family history of hypertension?
- Has Dr X adequately explained his opinion that the study of 850 stroke patients supports a conclusion that in C's case the chance of recovery with timely intervention would have been more than doubled?
- Are the Bradford Hill guidelines relevant to this problem? See [1.5 Correlation and causation](#). How do you apply them?
- Do you detect any other statistical problems?

Further reading

Sheila M Bird and Jane L Hutton, '[Statistician expert witnesses in agreement on relative hazards](#)'. (2012) 18(2) Clinical Risk 58 discusses a case where experts agreed on what they could and could not say about a case involving a patient who experienced pulmonary embolism after admission to Accident and Emergency, and who had not been prescribed a particular drug.

3.5 Case study three – Family

You are acting for WD, the mother of K aged 17 months. She is opposing an application by the local authority for a care and placement order. One of the issues is whether she is addicted to or abuses alcohol. Technical tests on samples of her hair have been carried out by experts which are capable of detecting recent alcohol consumption.

You are preparing for the pre-trial meeting of expert witnesses, with Dr Bunting of EZHairTest and Mr Temple from Sandringham Forensic Laboratories. What are the questions you should ask at the meeting?

Harlshire County Council ('the local authority') applies to the court for a care order and a placement order in respect of K. K is aged 17 months. WD ('the mother') opposes the local authority's applications.

K has some physical, behavioural and developmental difficulties. The local authority is concerned about the current and future capability of the mother to look after K. The mother vehemently opposes this. The mother has bipolar disorder, Obsessive Compulsive Disorder, and has previously also been hospitalised for excess alcohol consumption on at least three occasions. The mother is single. The father is unknown.

There were concerns that in at least the first and second trimester of the pregnancy the mother had abused alcohol and that she had also been taking medication for her mental health problems.

WD is undergoing a Care Programme Approach ('CPA'). WD is currently in the care of an NHS psychiatrist, Dr A. Dr A is submitting evidence that WD, the mother, is now undergoing therapy and support, including a programme of medication, to the effect that sustained progress has been made since K was born. Dr A's evidence suggests that the mother may be able to care for K even while still undergoing treatment. Dr A reports that her bipolar disorder has had severe episodes of mania, which are being addressed through the medication, therapy, and support. Dr A states that having a low level of alcohol consumption, ideally abstinence, is required to ensure that the treatment is successful. Alcohol consumption could trigger or worsen bipolar symptoms and reduce the effectiveness of the medication. Chronic or binge drinking could be particularly detrimental to WD's health and function.

The mother states that she drank heavily until around a year and a half before the baby was born, and she had been drinking a mixture of wine and spirits each day.

The mother agreed that she had previously been hospitalised due to alcoholism, and agreed to attend an alcohol avoidance project, CCA.

On her first visit to CCA she was breathalysed and showed an alcohol reading of 165 mg per 100 ml. She missed four out of eight subsequent appointments but did give negative breath tests when she did attend.

The mother states that she is trying hard to keep her alcohol intake low, does not keep alcohol in the house, and has tried to avoid drinks when out with friends.

She agreed to two separate hair strand tests for alcohol consumption and a blood test.

The Society of Hair Testing (SOHT) recommends testing for the ethanol metabolites ethyl glucuronide (EtG) and fatty acid ethyl esters (FAEEs), each of which can be measured in hair as direct markers of alcohol consumption. The SOHT notes that treatments such as bleaching, perming and dying may lead to lower measurements of EtG or false negative results, and may also influence concentrations of FAEEs. The SOHT also states that EtG appears not to be influenced by hair care products but hair care products that contain ethanol, such as some hairsprays or hair lotions may lead to false positive FAEEs.

In humans, head hair grows one centimetre (cm) per month on average, with a range of between approximately 0.7 and 1.5 cm.

EZHairTest conducted tests on two occasions. Dr Bunting from EZHairTest has supplied a report which states that the first sample taken six weeks ago (4cm, from close to the roots) had produced a result of EtG = 6.9 pg/mg and FAEEs = 0.31 ng/mg. Dr Bunting suspected that the hair showed signs of artificial colouring and treatment, and arranged for a second test. A second sample taken three weeks ago (6cm, from close to the roots) found EtG = 5pg/mg and FAEEs = 0.13 ng/mg.

The blood test was for Phosphatidylethanol (PEth) and was conducted at the same time as the second hair sample test by EZHairTest. Dr Bunting's report states that PEth is only produced when a person has consumed alcohol, with a detection period of up to 30 days (3-4 weeks) and a sensitivity rate (a true positive rate) of over 99%, with a false positive rate of less than 1%. Dr Bunting stated that over 20ng/ml is evidence of excessive alcohol abuse, with 100ng/ml or above strong evidence of heavy binge drinking. The PEth result for WD was 12ng/ml.

The local authority requested a separate set of tests from Sandringham Forensic Laboratories. They conducted a hair sample test. Mr Temple has supplied a report which states that at the request of

the local authority's solicitors, they took a 7.5cm sample from the root of WD's scalp, and found a FAEs reading of 0.21 ng/mg.

The local authority plans to use these results in its application to the court.

The Society of Hair Testing's 2016 Consensus for the Use of Alcohol Markers in Hair for Assessment of both Abstinence and Chronic Excessive Alcohol Consumption¹³ includes statements to the effect that:

- Hair taken from the vertex region of the scalp is preferred for testing.
- Tests for both EtG and FAEs may be affected by cosmetic treatments and thermal hair straightening tools.
- EtG tests can be affected by bleaching, perming and dying of hair, including potential for false negative results. Hair care products, including those that contain ethanol, do not appear to affect EtG.
- Concentrations of FAEs may be influenced by bleaching, perming, colouring of hair, and hair care products that contain ethanol may lead to false positive test results for FAEs.
- For abstinence testing, EtG is preferred to FAEs (details below). For chronic excessive alcohol consumption, either or both of EtG and FAEs can be used.
- Interpretation of EtG: For abstinence testing, less than 7 pg/mg "does not contradict self-reported abstinence" for the period before sampling; greater than or equal to 7 pg/mg in proximal scalp hair up to 6 cm "strongly suggests repeated alcohol consumption", even with a negative FAEs result.¹⁴ For chronic excessive alcohol consumption, EtG of >30 pg/mg in the proximal scalp hair up to 6cm "strongly suggests" such consumption. In all cases, results from samples less than 3cm or greater than 6cm long should be interpreted with caution.
- FAEs: Using FAEs alone is not recommended for assessing abstinence. When false negative EtG results are suspected in abstinence testing, a negative result for FAEs is a value up to 0.12 ng/mg for 0-3 cm proximal scalp hair, or up to 0.15 ng/mg for 0-6cm proximal scalp hair. A positive FAE result with an EtG measurement less than 7 pg/mg result "does not clearly disprove abstinence, but indicates the need for further monitoring".

¹³ Society of Hair Testing, '[2016 Consensus for the Use of Alcohol Markers in Hair for Assessment of both Abstinence and Chronic Excessive Alcohol Consumption](http://www.soht.org/images/pdf/Revision%202016_Alcoholmarkers.pdf)' (Society of Hair Testing, 2016) <www.soht.org/images/pdf/Revision%202016_Alcoholmarkers.pdf> accessed 6 October 2017

¹⁴ Note that this is not presented in a form recommended by ENFSI Guidelines

For chronic excessive alcohol consumption, results of greater than 0.35 ng/mg for ethyl palmitate in 0–3 cm proximal scalp hair, or greater than 0.45 ng/mg in 0 – 6cm scalp hair, is considered a positive result. Caution is recommended for hair from other than the scalp, or from other lengths of hair.

Points for discussion

- What questions might you ask of the test results for alcohol?
- Would you want expert comment on either or both test results? Consider the time and type of hair samples taken and issues with testing and sampling. Refer to [Refresher: The statistician's toolbox](#), [1.6 False positives and false negatives](#), [1.8 Expert opinion evidence](#) and [2.3 The scientific method](#).
- What relevance does the timescale of WD's activities have to the tests and the results?
- Is WD expected to maintain abstinence or reduce her alcohol intake?

The national standards for expert witnesses in family courts is given in Family Court Practice Direction 25B.

3.6 Case study four – Planning

You are instructed by RASMO (Residents Against Supermarkets in Marsh town Organisation) to represent it at the following planning inquiry, to cross-examine Retailer Developments PLC's (RD) highway consultants, and to present the case for RASMO on transportation issues. You are due to meet their expert witness Professor Whitehead in conference. RASMO does not wish to instruct a transport consultant, on grounds of cost, but wishes Professor Whitehead to give evidence.

Retailer Developments PLC (RD) wants to build a new superstore in an edge-of-centre location on a greenfield site at Marsh town, a market town in the Cotswolds. The town has a narrow high street (the B9990) possessing some ten local shops on each side of the street and a supermarket owned by a local trader, with 1,200 sq. m. of floorspace and a small car park at the rear. There are 120 homes within 10 minutes' walk to the High Street ('the 10-minute-walk isochrone'), 400 homes within a 20-minute-drive isochrone, and 600 more homes within a 30-minute-drive isochrone. On a neutral weekday such as Wednesday or Thursday the supermarket attracts about 700 shoppers between 8.30 am and 6.00 pm.

There is a small primary school in Church Street which caters for 120 pupils, half of whom are brought to school by car (figures supplied by the school secretary).

It is believed that the existing car parking arrangements in the High Street deter many of those living within the 20-minute-driving and 30-minute-driving isochrones to shop in the High Street. Surveys were carried out over two weeks by local volunteers who were seated at the entrance to the supermarket. Their overall results showed that less than 10 % of the shoppers have travelled more than 30 minutes to get there.

Results of the survey

Distance travelled	Walked 10 minutes or less	Drove 20 minutes or less	Drove between 20 & 30 minutes	Drove over 30 minutes	Total
Number	180	320	140	60	700
Percentage of total	26 %	46 %	20 %	8 %	100 %

The inhabitants of Marsh town have nevertheless been concerned about traffic congestion in the town centre over recent years.

The site for the proposed superstore is on the western edge of the town, also within the High Street's 10-minute-walk isochrone. It will be a large store with a floorspace of 5,000 sq m. It is part of RD's supporting planning statement that a store of this size will significantly improve the retail offer presently available to current shoppers in the High Street. It is also predicted to attract more shoppers from further afield, both inside and outside the 30-minute-drive isochrones. For many the only alternative is a one-hour drive into Wibchester, some 20 miles away to the east where, on the fringes, there are overlapping drive time catchments. There is no competing store on the western side of Marshtown. The store will therefore be particularly attractive to shoppers both within and outside the 30-minute-drive isochrones from that side.

The new superstore will have extensive free car parking at ground level. RD claims that this will relieve traffic congestion within the town.

To maximise its customer appeal, RD also wants the proposed development to have a petrol filling station (PFS). This will be advertised on the approaches to the slip road leading to Marshtown from the A995 dual carriageway, which is one mile away from the High Street. There are no PFSs on that road for 16 miles in either direction.

It is also part of the scheme that there will be pedestrian walks to the town centre, so that when people have carried out their shopping at the new store they will be able to walk in and carry out linked shopping trips in the High Street. It is claimed that this will reduce the vehicular traffic impact on the town centre.

There is a lot of local opposition to the proposed superstore. The shopkeepers with units on the High Street in Marshtown are convinced that the superstore, because of its highly competitive pricing policies, will damage their trade. The local preservation society perceives that a store in the proposed location will detract from the historic charms of Marshtown. The Campaign to Protect Rural England are concerned that the store location on a greenfield site is contrary to government planning policy.

A significant objection comes from a local residents' group which operates under the acronym, 'RASMO', the Residents Against Supermarkets in Marshtown Organisation. RASMO is concerned about the traffic likely to be induced into the town by the new development. RASMO is led by Professor William Whitehead, a local resident and retired professor of mathematics. He has calculated that, having regard to (1) the number of car-borne shoppers who currently shop in the High Street; (2) the number of school trips which are made by car; (3) the estimated number of additional car-borne shopping trips to Marshtown which will be generated by the new superstore; and (4) the estimated number of trips which will be made exclusively to the PFS, the B9990 passing

through the town will be in a permanent state of gridlock on neutral Wednesdays and Thursdays throughout the year.

Existing supermarket 1,500 sq. m with 700 shoppers on a neutral day

New supermarket @ 5,000 sq. m estimated to serve 2,500 shoppers:

Transfers from High Street:

10 minute walkers transferring from the High Street:	80
Drivers travelling for less than 20 minutes transferring:	250
Drivers travelling between 20 and 30 mins transferring:	100
Drivers travelling for more than 30 minutes transferring	50
Total transferred trips:	480

New shoppers (additional trips) induced by new supermarket = 2,020 (calculated from: 2,500 – 480)

(Any separate trips to the PFS unassociated with shopping will be additional but have not been calculated.)

Professor Whitehead also wants to present evidence that the projected increase in traffic flows will increase the level of pollution on the High Street above allowable guidelines.

Points for discussion

- What are the anticipated lines of attack on Professor Whitehead? Refer to the Guidelines above. How do you deal with potential attacks?
- Do you foresee problems with the survey taken at the door to the supermarket? Refer to [Refresher: The statistician's toolbox, 2.2 Inferential statistics and probability](#) and [2.3 The scientific method](#). How do you address any problems?
- Examine Professor Whitehead's estimates (1) (2) (3) and (4). How reliable are they? Is there any further research you would recommend to reinforce them?

SECTION 4: CURRENT AND FUTURE ISSUES

This brief guide can provide only a taster of the issues surrounding statistics, probability and law.

This section begins with a note on current activities on expert evidence, including in the areas of statistics, probability and law.

We conclude with a brief summary of four areas which are controversial in some way, or currently under debate. These are worth being aware of, to avoid, for example, your expert witness attempting to present evidence in a form that is currently not permitted due to existing case law. As these are live issues in law, they may be subject to change. Updates will be made in a subsequent edition of this guide.

Activities improving the reliability of expert evidence

In its 2011 report, the Law Commission recommended the introduction of a new statutory regime for assessing the reliability of expert evidence. Instead, amendments to the Criminal Procedure Rules and the Criminal Practice Direction were made by the Government and the Lord Chief Justice, respectively. The Government has not given any indication that it intends to revisit that report. The formal responses to the Law Commission's reforms may now be considered to be complete.

However, efforts to improve the reliability and assessment of expert evidence continue through a number of projects and activities, of which you should be aware. The Lord Chief Justice, the Royal Society, and the Royal Society of Edinburgh have established a working group to create primers for use by the judiciary, practitioners and juries in explaining and understanding forensic science when used in court. The Forensic Science Regulator is seeking to establish a quality standard for interpretation of evidence. The Leverhulme Research Centre for Forensic Science at the University of Dundee has been established to raise the standards of forensic science used in courts. Legal bodies, including the ICCA, are also developing training and professional development programmes, and initiatives advancing understanding of statistics, probability and law are underway in several academic settings.

Further activity may be required to improve the quality of scientific evidence to the degree identified as necessary by the Law Commission. In addition, the drive to improve the standard of expert evidence has focused mainly on criminal litigation; the need for reform is also present in the civil law context, where there has been less activity to date.

The likelihood ratio

The likelihood ratio is a measure of the value of the evidence and is used to help determine which, if either, of the competing propositions in a trial is true. The likelihood ratio can be set out in probabilistic terms as being the ratio of two conditional probabilities. In general terms, and set at the level of the offence (as opposed to activity or source level)¹⁵, these are:

- A. The probability of observing the evidence *given* that the defendant is liable (prosecution proposition); relative to:
- B. The probability of observing the evidence *given* that the defendant is not liable (defence proposition).

Either probability would be inadequate and misleading by itself. As the RSS (Aitken *et al.* 2010) has said:

Even if the evidence is unlikely assuming innocence, it could conceivably be even more unlikely assuming guilt. The probative value of the evidence cannot be assessed by examining only one of the two competing propositions.

Currently, this approach is supported within forensic science and set out within the ENFSI *Guideline for Evaluative Reporting in Forensic Science*.

The Law Commission's 2011 report, *Expert Evidence in Criminal Proceedings in England and Wales*, recommended an expansion of the approach of two competing propositions to more forms of evidence. They recommended:

*where an expert witness is called by a party to give a reasoned opinion on the likelihood of an item of evidence under a proposition advanced by that party, the expert's report must also include, where feasible, a reasoned opinion on the likelihood of the item of evidence under one or more alternative propositions (including any proposition advanced by the opposing party)*¹⁶

They argued for this on the basis that all expert witnesses have an overriding duty to provide impartial evidence. They noted that it may not always be feasible to provide such an alternative.

This requirement is **not** present in the current (2015) Criminal Procedure Rules or the Criminal Practice Direction 19A.

¹⁵ Propositions in a trial could be at source level, sub-source level, activity level or offence level. See section 3, particularly paragraphs 3.4 – 3.8, of *Fundamentals of probability and statistical evidence in criminal proceedings* (Aitken *et al.* 2010) for definitions and discussion.

¹⁶ Law Commission, *Expert Evidence in Criminal Proceedings in England and Wales* (Law Com No. 325, 2011) para 7.21 (2)(c)

The *ENFSI Guideline for Evaluative Reporting in Forensic Science* recommends the use of likelihood ratios to provide a balanced approach. The guidelines also recommend that the reporting be transparent. The case of *R v T [2010]* demonstrated the importance of an expert evaluating a likelihood ratio in a transparent and justifiable way.

In the case of *R v T [2010]*, T appealed a murder (case redacted) on the basis of identification. Footwear marks left at the crime scene had been used to identify a type of trainer that T wore. The forensic expert had used a likelihood ratio including assessments of the pattern, size, wear and damage to the shoe seen in the shoe print to conclude that "there is at this stage a moderate degree of scientific evidence to support the view that the [Nike trainers recovered from the appellant] had made the footwear marks" (paragraph 24 of the judgement). These calculations, and the assumptions within them, were not evident in the expert witness' report. The court disagreed with this use of likelihood ratios as the numerical values used in the calculations were not considered reliable enough – statistical information on shoes in the UK was not available in sufficient detail. The court ruled that footwear mark comparison should be limited to the expression of non-probabilistic evaluative opinions by experts. The conviction was quashed.

R v T also determined that the forensic scientists should not use data in their formulation of their expert statements without making the source of the data explicit, and should use only data that were appropriate for the task.¹⁷

Bayesian reasoning

Bayesian reasoning refines the likelihood ratio by using the likelihood ratio multiplied by **prior odds** – the odds of something being the case before the introduction of new evidence – to develop the **posterior odds** – the odds of something being the case in the light of the new evidence.

Statisticians consider likelihood ratios to be a component of Bayesian reasoning. However, though the likelihood ratio approach to evaluating evidence is supported within forensic science, and set out within the *ENFSI Guideline for Evaluative Reporting in Forensic Science*, the role of Bayesian reasoning within the courts is less clear-cut, and controversial.

¹⁷ *R v South [2011]* is an example of a case where footmark evidence was permitted under the *R v T* guidelines. See also Redmayne *et al.* (2011) and Robertson *et al.* (2011) for critique of the *R v T* judgement.

Introductions to Bayesian reasoning are provided in the RSS guide *Fundamentals of probability and statistical evidence in criminal proceedings* (Aitken et al. 2010, paragraphs 2.25 – 2.31) and *Bayes and the Law* (Fenton et al., 2016), which includes a review of cases using Bayesian reasoning.

Many statisticians and scientists consider Bayesian reasoning, and related tools such as Bayes' formula and Bayesian nets, to simply be an extension of logical reasoning, a heuristic tool for inferential reasoning. Many statisticians and mathematicians believe that Bayesian reasoning should be used by experts in formulating their expert judgments, but there is not a consensus on the use of Bayesian reasoning in court for either criminal or civil cases.

Denis Adams was tried for the rape of a woman in a park. Evidence in his favour included the victim's declaration at an identity parade that Adams was not her attacker, and an alibi from his girlfriend. A DNA profile of evidence from the scene matched that of a DNA profile of Adams, which the prosecution said had a random match probability of 1 in 200 million. At trial, a jury found him guilty. Two subsequent appeals sought to address the interpretation of the 1 in 200 million figure in the context of the evidence in favour of Adams' innocence, by guiding the jury through Bayesian reasoning. Both appeals failed, with the court each time stating an objection to the use of Bayesian reasoning by juries to evaluate non-scientific evidence.

In *R v T*, the court drew upon the judgement of *R v Adams* to conclude that “*outside the field of DNA (and possibly other areas where there is a firm statistical base), [...] Bayes Theorem and likelihood ratios should not be used.*” (paragraph 90). The court made no attempt to evaluate the use of likelihood ratios in relation to other types of scientific evidence.

The position remains unclear. Any change to the position would be particularly dependent on how this approach is presented, or the extent to which it can be made understandable to lawyers and juries.

The use of epidemiological evidence and the ‘doubling of risk’ test

As we saw in [1.5 Correlation and causation](#), techniques drawn from epidemiology could be appropriate for addressing issues of causation. We have also noted the tension between the legal approach focused on the causes of effects, and the scientific approach focused on the effects of causes (see [Differences of approach](#) in [3.1 Communicating with experts about statistical evidence](#)).

As Lord Nicholls said in *Gregg v Scott*:

Statistical evidence, however, is not strictly a guide to what would have happened in one particular case. Statistics record retrospectively what happened to other patients in more or less comparable situations. They reveal trends of outcome. They are general in nature. The different

*way other patients responded in a similar position says nothing about how the claimant would have responded. Statistics do not show whether the claimant patient would have conformed to the trend or been an exception from it. They are an imperfect means of assessing outcomes even of groups of patients undergoing treatment, let alone a means of providing an accurate assessment of the position of one individual patient.*¹⁸

Legal and statistical scholars argue that the court's approach to epidemiological evidence is not consistent or sufficient (for example, Dawid *et al.*, McIvor, Turton).

In England and Wales, these issues are particularly noticeable where a court is to determine what caused a disease where there are competing explanations; for example, smoking, asbestos or something else.¹⁹ Where there are alternative explanations, the courts may be persuaded to accept a causative link on the balance of probabilities based on statistical or expert evidence where exposure to a relevant chemical or agent can be shown to have at least doubled the risk which exists in the absence of such exposure of contracting that disease.²⁰ This approach is controversial though it has succeeded, for example, in *Novartis Grimsby Ltd v Cookson*, where Cookson successfully argued that exposure to workplace chemicals, rather than his past smoking habit, had led to his bladder cancer.²¹

In *Heneghan v Manchester Dry Docks Ltd* [2016] EWCA Civ 86; [2016] 1 W.L.R. 2036, Lord Dyson MR said:

The 'doubles the risk' test is one that applies epidemiological data to determining causation on the balance of probabilities where medical science does not permit determination with certainty of how an injury was caused. If statistical evidence shows that a tortfeasor more than doubled the risk that the victim would suffer the injury, it follows that it is more likely than not that the tortfeasor caused the injury.

The discussion in *Sienkiewicz v Greif (UK) Ltd* [2011] 2 A.C. 229 of this method of assigning causation shows that this approach needs to be treated both carefully and also in context.²² In spite of strong criticism from some legal scholars (see, for example, McIvor, Turton), courts continue to apply a 'doubles the risk' test. The courts' view on doubling of risk and causation is likely to develop as

¹⁸ [2005] UKHL 2; [2005] 2 A.C. 176 [28]

¹⁹ See *Jones v Secretary of State for Energy and Climate Change* [2013] EWHC 1023 (QB)

²⁰ This is sometimes referred to as a relative risk being greater than 2: see [1.4 Absolute and relative risk](#), though this is a convenient shorthand rather than a reliable rule: it implies that more than one-half of the cases can be attributed to exposure, but makes assumptions about the individuals in the instant case, and epidemiologists do not treat findings of a relative risk greater than two as proof of causation.

²¹ [2007] EWCA Civ 1261. See also *Jones v Secretary of State for Energy and Climate Change* [2012] EWHC 2936 (QB) (the Phurnacite Workers Group Litigation)

²² See particularly Lord Phillips in paragraphs 94 to 96, doubting its applicability to mesothelioma cases, and Lord Rodger in paragraph 149 of *Sienkiewicz v Greif (UK) Ltd* [2011] 2 A.C. 229

medical, scholarly and legal thinking develop. An American judge said in *Merrell Dow Pharmaceuticals Inc v Havner (1997)*:

The use of scientifically reliable epidemiological studies and the requirement of more than a doubling of the risk strikes a balance between the needs of our legal system and the limits of science. We do not hold, however, that a relative risk of more than 2.0 is a litmus test or that a single epidemiological test is legally sufficient evidence of causation. Other factors must be considered. As already noted, epidemiological studies only show an association.

Estimates of life expectancy

Statistics are used in personal injury cases to determine life expectancy: the Government's actuarial or "Ogden" tables use UK mortality statistics to provide appropriate multipliers for calculating the net present value (NPV) of future loss or expense.

Sometimes general mortality tables are inappropriate because it is agreed that the claimant, usually as a result of the accident but sometimes because of unrelated illness or disease, has a much lower life expectancy than normal. Specialist statisticians have created their own databases derived, for example, from populations of those who have suffered traumatic brain injury from which a tailor-made life table for the individual claimant may be created.

This is also controversial when used in court because of the need to ensure that the population from which the data is derived is comparable to the claimant's situation. For example, if the claimant is going to receive sufficient damages to provide all the nursing care needed in future, he or she may be less at risk of premature death than those victims of accidents who have no claim for damages, and do not have access to such beneficial care.²³

References and further reading

A Philip Dawid, David L Faigman and Stephen E Fienberg, 'Fitting Science Into Legal Contexts' [2013] *Sociological Methods & Research* 359

Peter Donnelly, '[Appealing statistics](#)' 2(1) *Significance* 46 Statistician Peter Donnelly recounts his experience of being involved in the Adams case, commenting on the limits of the approach.

ENFSI, '[ENFSI guideline for evaluative reporting in forensic science](#)' (ENFSI 2015) <<http://enfsi.eu/news/enfsi-guideline-evaluative-reporting-forensic-science/>> accessed 6 October 2017

Norman Fenton, Martin Neil, and Daniel Berger, 'Bayes and the Law' 3 *Annual Review of Statistics and Its Application* 51

Law Commission, *Expert Evidence in Criminal Proceedings in England and Wales* (Law Com No. 325, 2011)

Merrell Dow Pharmaceuticals Inc v Havner (1997) 953 SW 2d 706 (Tex. 1997)

²³ *Robshaw v United Lincolnshire Hospitals NHS Trust* [2015] EWHC 923 (QB); [2015] Med.L.R.339 is an example of such a case, see paragraphs 107 to 132.

Claire McIvor, 'Debunking some judicial myths about epidemiology and its relevance to UK tort law' [2013] Medical Law Review 553

Novartis Grimsby Ltd v Cookson [2007] EWCA Civ 1261

R v Adams [1996] 2 Cr App R 467, [1996] Crim LR 898, CA and *R v Adams* [1998] 1 Cr App R 377

R v South [2011] EWCA Crim 754 CA – a case subsequent to and referencing *R v T*, where footmark evidence was permitted under the *R v T* guidelines

R. v T (footwear mark evidence) [2010] EWCA Crim 2439 [2011] 1 Cr. App. R. 9

Mike Redmayne, *Expert Evidence and Criminal Justice* (Oxford University Press 2001)

Mike Redmayne, Paul Roberts, Colin Aitken, Graham Jackson, 'Forensic science evidence in question' 5 (2011) Criminal Law Review 347 includes critique of the *R v T* judgement

Bernard Robertson, G A Vignaux and Charles E H Berger, 'Extending the confusion about Bayes' 74 (2011) Modern Law Review 430 includes critique of the *R v T* judgement

Colin Aitken, Paul Roberts, Graham Jackson, *Fundamentals of probability and statistical evidence in criminal proceedings* (Royal Statistical Society 2010)

Sienkiewicz v Greif (UK) Ltd [2011] UKSC 10; [2011] 2 AC 229

The correspondence in Science and Justice 54 (4) of Alex Biedermann, Tacha Hicks, Franco Taroni, Christophe Champod and Colin Aitken, 'On the use of the likelihood ratio for forensic evaluation: Response to Fenton et al.' 54 (2014) Science and Justice 316 and Norman Fenton, David Lagnado, Anne Hsu, Daniel Berger and Martin Neil, 'Response to "On the use of the likelihood ratio for forensic evaluation: Response to Fenton et al."' 54 (2014) Science and Justice 319 illustrates current discussion in the application of the likelihood ratio, and whether the two propositions need to be exhaustive or not.

Gemma Turton, *Evidential Uncertainty in Causation in Negligence* (Hart Publishing 2016)

UK Government's Actuary Department, *Actuarial Tables with explanatory notes for use in Personal Injury and Fatal Accident Cases* (7th edn, The Stationery Office 2011)

FURTHER RESOURCES

General statistics, probability and the scientific method

Sense About Science, '[Making Sense of Statistics](http://senseaboutscience.org/activities/making-sense-of-statistics/)' (Sense About Science, 29 April 2010)
<<http://senseaboutscience.org/activities/making-sense-of-statistics/>> is a short statistics primer.

National Research Council, [Reference Manual on Scientific Evidence: Third Edition](#) (The National Academies Press 2011). This book is American but its two sections on 'How Science Works' (p. 37-54) and 'Reference Guide on Statistics' (p.211-302) are very clear and include examples of legal application.

Huff D, *How to Lie with Statistics* (New edn, Penguin 1991). Whilst this book was first produced in 1954, it remains one of the highest selling books on statistics. It covers the misuse of statistics from biased samples, to dubious graphs and trends.

Blastland M and Dilnot A, *The Tiger That Isn't* (Profile Books 2008). A popular introduction to statistics and probability

Freeman J V, Walters S J and Campbell M J, *How to Display Data* (BMJ Books 2008). A short and accessible guide to presenting data.

Cancer Research UK provides some useful information about the scientific method on its website at Cancer Research UK, '[What is a randomised trial?](#)' (Cancer Research UK, 29 October 2014)
<www.cancerresearchuk.org/file/34246> accessed 6 October 2017.

Statistics, probability and the law

The RSS has published a set of four practitioner guides which looks at communicating and interpreting statistical evidence in the administration of criminal justice. They are intended to assist judges, lawyers, forensic scientists and other expert witnesses in coping with the demands of modern criminal litigation.

1. Colin Aitken, Paul Roberts, Graham Jackson, [Fundamentals of probability and statistical evidence in criminal proceedings](#) (Royal Statistical Society 2010)
2. Roberto Puch-Solis, Paul Roberts, Susan Pope, Colin Aitken, [Assessing the probative value of DNA evidence](#) (Royal Statistical Society 2012) Includes a short guide to DNA and testing procedures.
3. Paul Roberts and Colin Aitken, [The logic of forensic proof: inferential reasoning in criminal evidence and forensic science](#), (Royal Statistical Society 2014)
4. Graham Jackson, Colin Aitken and Paul Roberts, [Case assessment and interpretation of expert evidence](#) (Royal Statistical Society 2015)

The Royal Statistical Society's Statistics and Law section welcomes members of the statistical, legal or forensic scientific communities to join their meetings on topics of criminal and civil law www.rss.org.uk/law

Aitken C and Taroni F, 'Fundamentals of statistical evidence - a primer for legal professionals', (2008) 12 International Journal of Evidence and Proof 181

Dawid, A P, 'Statistics and the law', in Bell A, Swenson-Wright J and Tybjerg K (eds), *In Evidence* (Cambridge University Press 2008)

Kadane J B '*Statistics in the Law: A Practitioner's Guide, Cases, and Materials*' (Oxford University Press 2008). Whilst mostly based on American law, Part 5 of this book reviews English law and provides information on how statistical analysis can be shaped to address relevant legal questions.

Schneps L and Colmez C, [Math on Trial: How Numbers Get Used and Abused in the Courtroom](#) (Basic Books 2013). Mathematicians Leila Schneps and Coralie Colmez describe ten trials from the nineteenth century to today where maths has been misused including the cases of Sally Clark, Amanda Knox, and Lucia de Berk.

Finkelstein M O, *Basic Concepts of Probability and Statistics in the Law* (Springer Science 2009)

Gastwirth, J L *Statistical Reasoning in Law and Public Policy. Vol. 1 Statistical Concepts and Issues of Fairness. Vol. 2. Tort Law, Evidence and Health.* (Academic Press 1988) provides a US perspective.

UK legal commentary and guidance on statistics and probability

Law Commission, *Expert Evidence in Criminal Proceedings in England and Wales* (Law Com No. 325, 2011)

Judicial College (2017) *The Crown Court Compendium*. Relevant sections include DNA at Chapter 15-8.

Forensic science

Sense About Science and EUROFORGEN, 'Making Sense of Forensic Genetics' (*Sense About Science*, 25 January 2017) <<http://senseaboutscience.org/activities/making-sense-of-forensic-genetics/>> accessed 6 October 2017. An introductory guide to forensic science, focusing on the use of DNA in criminal investigations.

National Research Council, *Reference Manual on Scientific Evidence: Third Edition* (The National Academies Press 2011). Includes a Reference Guide on DNA Identification Evidence.

ENFSI, 'ENFSI guideline for evaluative reporting in forensic science' (ENFSI 2015) <<http://enfsi.eu/news/enfsi-guideline-evaluative-reporting-forensic-science/>> accessed 6 October 2017

'Probability and statistics in forensic science' was a scientific programme of the Isaac Newton Institute, Cambridge in 2016. Outputs from the programme, including talks and papers by lawyers, statisticians and forensic experts, are being published at <https://www.newton.ac.uk/event/fos>

INDEX OF STATISTICAL TERMS

These terms are defined in context and in-line in the text.

absolute risk (AR).....	16	median.....	9
analysing data.....	8	mode.....	9
Bayesian reasoning.....	67	model.....	38
Bradford Hill guidelines.....	22	non-sampling error.....	41
collecting data.....	8	outliers.....	9
conditional probability.....	41	population.....	8
confidence interval.....	39	posterior odds.....	67
confidence level.....	39	power.....	45
confounding variable.....	19	prior odds.....	67
convenience samples.....	33	probability.....	36
correlations.....	18	prosecutor's fallacy.....	25, 26
data.....	8	random match probability (RMP).....	31
defence fallacy.....	25, 28	randomised controlled trial.....	43
dependent.....	41	range.....	9
epidemiological studies.....	18	relative risk (RR).....	16
epidemiology.....	21	representative.....	8
estimate.....	8, 39	risk.....	16
event.....	14	sample.....	8
false negative.....	24	sample mean.....	39
false positive.....	24	sampling.....	33
frequency.....	9	sampling error.....	41
independence.....	13	significance.....	44
independent.....	13, 37	skew.....	9
interpreting data.....	8	source probability error.....	32
interquartile range (IQR).....	9	spurious correlation.....	19
interval estimate.....	39	standard deviation (SD).....	9
likelihood ratio.....	66	transposing the conditional.....	26
margin of error.....	40	variable.....	14
mean.....	9		

ACKNOWLEDGEMENTS

The ICCA and the RSS gratefully acknowledge the assistance in the preparation of this guide of the following individuals. These include members of the ICCA Expert Evidence in Proceedings group and the RSS Statistics and Law section.

For the ICCA:

Aadhithya Anbahan
David Bentley QC
Sarah Clarke QC
Adam Constable QC
Siân Gough
Martin Griffiths QC
HHJ David Hodge QC
Michael Kent QC
Alison Levitt QC
Colin McCaul QC
Professor David Ormerod QC
John Ross QC
Rebecca Stubbs QC
Derek Wood CBE QC
And the late Dame Frances Patterson DBE QC

For the Royal Statistical Society:

Professor Colin Aitken CStat
Professor Sandra Eldridge
Professor Joseph Gastwirth
Professor Jane Hutton CStat
Dr Graham Jackson
Scott Keir
Professor Kevin McConway CStat
Dr Claire McIvor
Professor Julia Mortera
Professor Niamh Nic Daeid
Dr Sue Pope
Dr Roberto Puch-Solis
Neil Sheldon CStat
Professor David Spiegelhalter CStat
Professor Cathal Walsh
Andrea Westall
Dr Amy Wilson

Copyright notice

- The Council of the Inns of Court (COIC) and the Royal Statistical Society (RSS) are the owners or licensees of all copyright in this guide. October 2017 - All rights reserved.
- You may read, print one copy or download this guide document for your own personal use.
- You **may not** make commercial use of this training document, adapt or copy it without our permission.
- Every effort has been made to acknowledge and obtain permission to use any content that may be the material of third parties. COIC will be glad to rectify any omissions at the earliest opportunity.
- Use of this guide document is subject to [our terms of use](#).

The Inns of Court College of Advocacy (ICCA)
9 Gray's Inn Square, London WC1R 5JD
United Kingdom
0207 822 0763

 www.icca.ac.uk

 [@Inns_College](https://twitter.com/Inns_College)

 info@icca.ac.uk

Part of the Council of the Inns of Court, Company Limited by
Guarantee – No. 8804708 Charity No. 1155640

The Royal Statistical Society
12 Errol Street, London EC1Y 8LX
United Kingdom
0207 638 8998

 www.rss.org.uk

 [@RoyalStatSoc](https://twitter.com/RoyalStatSoc)

 rss@rss.org.uk

The RSS is a registered charity No. 306096