Communicating and Interpreting Statistical Evidence in the Administration of Criminal Justice

1. Fundamentals of Probability and Statistical Evidence in Criminal Proceedings

Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses

Colin Aitken, Paul Roberts, Graham Jackson

ROYAL STATISTICAL SOCIETY

PRACTITIONER GUIDE NO 1

Fundamentals of Probability and Statistical Evidence in Criminal Proceedings

Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses

By

Colin Aitken, Professor of Forensic Statistics, University of Edinburgh,

Paul Roberts, Professor of Criminal Jurisprudence, University of Nottingham

Graham Jackson, Professor of Forensic Science, Abertay University

Prepared under the auspices of the

Royal Statistical Society's *Working Group on Statistics and the Law*

(Chairman: Colin Aitken)

Contents

Introduction to *Communicating and Interpreting Statistical Evidence in the Administration of Criminal Justice*

0.1 ***Context, Motivation and Objectives***

Statistical evidence and probabilistic reasoning today play an important and expanding role in criminal investigations, prosecutions and trials, not least in relation to forensic scientific evidence (including DNA) produced by expert witnesses. It is vital that everybody involved in criminal adjudication is able to comprehend and deal with probability and statistics appropriately. There is a long history and ample recent experience of misunderstandings relating to statistical information and probabilities which have contributed towards serious miscarriages of justice.

0.2 English and Scottish criminal adjudication is strongly wedded to the principle of lay fact-finding by juries and magistrates employing their ordinary common sense reasoning. Notwithstanding the unquestionable merits of lay involvement in criminal trials, it cannot be assumed that jurors or lay magistrates will have been equipped by their general education to cope with the forensic demands of statistics or probabilistic reasoning. This predictable deficit underscores the responsibilities of judges and lawyers, within the broader framework of adversarial litigation, to present statistical evidence and probabilities to fact-finders in as clear and comprehensible a fashion as possible. Yet legal professionals' grasp of statistics and probability may in fact be little better than the average juror's.

Perhaps somewhat more surprisingly, even forensic scientists and expert witnesses, whose evidence is typically the immediate source of statistics and probabilities presented in court, may also lack familiarity with relevant terminology, concepts and methods. Expert witnesses must satisfy the threshold legal test of competency before being allowed to testify or submit an expert report in legal proceedings.[1] However, it does not follow from the fact that the witness is a properly qualified expert in say, fingerprinting or ballistics or paediatric medicine, that the witness also has expert – or even rudimentary – knowledge of

[1] *R* v *Atkins* [2009] EWCA Crim 1876; *R* v *Stockwell* (1993) 97 Cr App R 260, CA; *R* v *Silverlock* [1894] 2 QB 766, CCR.

statistics and probability. Indeed, some of the most notorious recent miscarriages of justice involving statistical evidence have exposed errors by experts.

There is, in short, no group of professionals working today in the criminal courts that can afford to be complacent about its members' competence in statistical method and probabilistic reasoning.

0.3. Well-informed observers have for many decades been arguing the case for making basic training in probability and statistics an integral component of legal education (e.g. Kaye, 1984). But little tangible progress has been made. It is sometimes claimed that lawyers and the public at large fear anything connected with probability, statistics or mathematics in general, but irrational fears are plainly no excuse for ignorance in matters of such great practical importance. More likely, busy practitioners lack the time and opportunities to fill in persistent gaps in their professional training. Others may be unaware of their lack of knowledge, or believe that they understand but do so only imperfectly ("a little learning is a dang'rous thing"[2]).

0.4. If a broad programme of education for lawyers and other forensic practitioners is needed, in what should this consist and how should it be delivered? It would surely be misguided and a wasted effort to attempt to turn every lawyer, judge and expert witness (let alone every juror) into a professor of statistics. Rather, the objective should be to equip forensic practitioners to become ***responsible producers*** and ***discerning consumers*** of statistics and confident exponents of elementary probabilistic reasoning. It is a question of each participant in criminal proceedings being able to grasp at least enough to perform their respective allotted roles effectively in the interests of justice.

For the few legal cases demanding advanced statistical expertise, appropriately qualified statisticians can be instructed as expert witnesses in the normal way. For the rest, lawyers need to understand enough to be able to question the use made of statistics or probabilities and to probe the strengths and expose any weaknesses in the evidence presented to the court; judges need to understand enough to direct jurors clearly and effectively on the statistical or probabilistic aspects of the case; and expert witnesses need to understand

[2] Alexander Pope, *An Essay on Criticism* (1711).

enough to be able to satisfy themselves that the content and quality of their evidence is commensurate with their professional status and, no less importantly, with an expert witness's duties to the court and to justice.[3]

0.5 There are doubtless many ways in which these pressing educational needs might be met, and the range of possibilities is by no means mutually exclusive. Of course, design and regulation of professional education are primarily matters to be determined by the relevant professional bodies. However, in specialist matters requiring expertise beyond the traditional legal curriculum it would seem sensible for authoritative practitioner guidance to form a central plank of any proposed educational package. This would ideally be developed in conjunction with, if not directly under the auspices of, the relevant professional bodies and education providers.

The US Federal Judicial Center's *Reference Manual on Scientific Evidence* (2nd edn, 2000) provides a valuable and instructive template. Written with the needs of a legal (primarily, judicial) audience in mind, it covers a range of related topics, including: data collection, data presentation, base rates, comparisons, inference, association and causation, multiple regression, survey research, epidemiology and DNA evidence. There is currently no remotely comparable UK publication specifically addressing statistical evidence and probabilistic reasoning in criminal proceedings in England and Wales, Scotland and Northern Ireland.

0.6 In association with the Royal Statistical Society (RSS) and with the support of the Nuffield Foundation, we aim to fill this apparent gap in UK forensic practitioner guidance. This is the first of four planned *Practitioner Guides* on aspects of statistical evidence and probabilistic reasoning, intended to assist judges, lawyers, forensic scientists and other expert witnesses in coping with the demands of modern criminal litigation. The *Guides* are being written by a multidisciplinary team comprising a statistician (Aitken), an academic lawyer (Roberts), and two forensic scientists (Jackson and Puch-Solis). They are produced under the auspices of the RSS's Working Group on Statistics and the Law, whose membership includes representatives from the judiciary, the English Bar, the Scottish

[3] *R* v *B(T)* [2006] 2 Cr App R 3, [2006] EWCA Crim 417, [176]. And see CrimPR 2010, Rule 33.2: 'Expert's duty to the court', reproduced in Appendix B, below.

Faculty of Advocates, the Crown Prosecution Service, the National Police Improvement Agency (NPIA) and the Forensic Science Service, as well as academic lawyers, statisticians and forensic scientists.

0.7 ***Users' Guide to this Guide – Some Caveats and Disclaimers***

Guide No 1 is designed as a general introduction to the role of probability and statistics in criminal proceedings, a kind of *vade mecum* for the perplexed forensic traveller; or possibly, 'Everything you ever wanted to know about probability in criminal litigation but were too afraid to ask'. It explains basic terminology and concepts, illustrates various forensic applications of probability, and draws attention to common reasoning errors ('traps for the unwary'). A further three *Guides* will be produced over the next three years. Building on the foundations laid by *Guide No 1*, they will address the following more discrete topics in greater detail: (2) DNA profiling evidence; (3) networks for structuring evidence; and (4) case assessment and interpretation. Each of these topics is of major importance in its own right. Their deeper exploration will also serve to elucidate and exemplify the general themes, concepts and issues in the communication and interpretation of statistical evidence and probabilistic reasoning in the administration of criminal justice which are introduced in the following pages.

0.8 This *Guide* develops a logical narrative in which each section builds on those which precede it, starting with basic issues of terminology and concepts and then guiding the reader through a range of more challenging topics. The *Guide* could be read from start to finish as a reasonably comprehensive primer on statistics and probabilistic reasoning in criminal proceedings. Perhaps some readers will adopt this approach. However, we recognise that many busy practitioners will have neither the time nor the desire to plough through the next eighty-odd pages in their entirety. So the *Guide* is also intended to serve as a sequence of self-standing introductions to particular topics, issues or problems, which the reader can dip in and out of as time and necessity direct. Together with the four appendices attached to this *Guide*, we hope that this modular format will meet the practical needs of judges, lawyers and forensic scientists for a handy work of reference that can be consulted, possibly repeatedly, whenever particular probability-related issues arise during the course of their work.

0.9 We should flag up at the outset certain challenges which beset the production of this kind of *Guide*, not least because it is likely that we have failed to overcome them entirely satisfactorily.

First, we have attempted to address multiple professional audiences. Insofar as there is a core of knowledge, skills and resources pertaining to statistical evidence and probabilistic reasoning which is equally relevant for trial judges, lawyers and forensic scientists and other expert witnesses involved in criminal proceedings, it is entirely appropriate and convenient to pitch the discussion at this generic level. The successful integration of statistics and probabilistic reasoning into the administration of criminal justice is likely to be facilitated if participants in the process are better able to understand other professional groups' perspectives, assumptions, concerns and objectives. For example, lawyers might be able to improve the way they instruct experts and lead their evidence in court by gaining insight into forensic scientists' thinking about probability and statistics; whilst forensic scientists, for their part, may become more proficient as expert witnesses by gaining a better appreciation of lawyers' understandings and expectations of expert evidence, in particular regarding the salience and implications of its probabilistic character.

We recognise, nonetheless, that certain parts of the following discussion may be of greater interest and practical utility to some criminal justice professionals than to others. This is another reason why readers might prefer to treat the following exposition and its appendices more like a work of reference than a monograph. Our hope is that judges, lawyers and forensic scientists will be able to extrapolate from the common core of mathematical precepts and their forensic applications and adapt this generic information to the particular demands of their own professional role in criminal proceedings. For example, we hope to have supplied useful information that might inform the way in which a trial judge might assess the admissibility of expert evidence incorporating a probabilistic component or direct a jury in relation to statistical evidence but we have stopped well short of presuming to specify formal criteria of legal admissibility or to formulate concrete guidance that trial judges might repeat to juries. We have neither the competence nor the authority to made detailed recommendations on the law and practice of criminal procedure.

0.10 The following exposition is also generic in a second sense directly related to the preceding observations. We hope that this *Guide* will be widely used in all of the United Kingdom's legal jurisdictions. It goes without saying that the laws of probability, unlike the laws of the land, are valid irrespective of geography. It would be artificial and sometimes misleading when describing criminal litigation to avoid any reference whatsoever to legal precepts and doctrines, and we have not hesitated to mention legal rules where the context demands it. However, we have endeavoured to keep such references fairly general and non-technical – for example, by referring in gross to "the hearsay prohibition" whilst skating over jurisdictionally-specific doctrinal variations with no bearing on probability or statistics. Likewise, references to points of comparative law – such as Scots law's distinctive corroboration requirement – will be few and brief. Readers should not expect to find a primer on criminal procedure in the following pages.

0.11 A third caveat relates to the nature of the information about probability and statistics that this *Guide* does contain, and it is possibly the most significant and difficult to articulate clearly. Crudely stated, the question is: how accurate is this *Guide*?

Insofar as accuracy is a function of detail and precision, this *Guide* cannot be as accurate as a textbook on mathematics or forensic statistics. The market is already well-served by such publications.[4] This *Guide* necessarily trades a measure of accuracy *qua* comprehensiveness for greater comprehensibility and practical usefulness, with references and further reading listed in the Appendices for those seeking more rigorous and exhaustive treatments. Our focus will be on the fundamentals of statistical evidence and probabilistic reasoning – and the generalisations contained in parts of this *Guide* are presented as mathematically valid generalisations.

Conversely, this *Guide* grapples with some conceptually difficult and intellectually challenging topics, aspects of which need to be expressed through specialist terminology and notation. Appendix A provides a glossary of such technical terms, which appear in the main text in ***bold italic.*** As with the law, we are assuming a non-specialist audience and have endeavoured to keep mathematical technicalities to a minimum. That said, it is perhaps worth stating at the outset that readers should not expect the following simplified

[4] See e.g. Aitken and Taroni (2004); Robertson and Vignaux (1995).

account of statistical evidence and probabilistic reasoning in criminal proceedings to be in any way *simplistic* or even simple to grasp in every respect. We take ourselves to be addressing a rather rarefied class of "general reader", comprised of criminal justice professionals who have a strong occupational interest, and indeed professional duty, to acquaint themselves with the fundamentals of probability and statistics and their implications for the routine conduct of criminal litigation.

0.12 "Accuracy", then, is partly a question of objective facts and partly a function of striking an appropriate balance for the purposes at hand between tractable generalisations and exhaustive technical detail. It is also a matter of irreducible *controversy*. Since scientific facts are popularly regarded as straightforwardly true or false, this observation requires elucidation.

Assuming the basic axioms of mathematics, mathematical propositions, theorems and solutions are either true or false, deductively valid or invalid. Likewise probabilistic calculations are either correct or incorrect. However, like any field of scientific inquiry, there remain areas of theory and practice that are subject to uncertainty and competing interpretations by specialists. Moreover, even if a particular mathematical result is undeniably sound, its potential *forensic applications* (including the threshold question of whether it should have any at all) may be matters of on-going debate and even intense controversy between proponents and their critics, who may be adopting different starting points and assumptions.

The following exposition is intended to present "just the essential facts" about statistical evidence and probabilistic reasoning in as neutral a fashion as possible. The specific issues, formulae, calculations and illustrations we present are meant to function as a kind of intellectual toolkit. We attempt to identify and explain the strengths and weaknesses of each tool without necessarily recommending its use for a particular forensic job. Whether or not readers already do or might in future choose to employ some of these tools in their own professional practice, we hope that this *Guide* will better equip readers to respond appropriately and effectively when they encounter other lawyers or scientists freely exploiting the statistics and probability toolkit in the course of criminal proceedings.

Where we occasionally deemed it impossible or inappropriate to steer clear of all controversy, we have endeavoured to indicate the range of alternative approaches and their respective merits. For the avoidance of any doubt, this *Guide* does not pursue any strategic or broader reformist objective, beyond our stated aim of improving the communication and interpretation of statistical evidence and probabilistic reasoning in the administration of criminal justice.

0.13 This *Guide* has evolved through countless drafts over a period of several years. It has benefited immeasurably from the generous (unpaid) input of fellow members of the RSS's Working Group on Statistics and the Law and from the guidance of our distinguished international advisory panel. The *Guide* also incorporates helpful suggestions and advice received from many academic colleagues, forensic practitioners, representative bodies and other relevant stakeholders. We are grateful in particular to His Honour Judge John Phillips, Director of the Judicial Studies Board, for his advice in relation to criminal litigation in England and Wales, and to Sheriff John Horsburgh who performed a similar advisory role in relation to Scottish law and practice. Whilst we gratefully acknowledge our intellectual debts to this extraordinarily well-qualified group of supporters and friendly critics, the time-honoured academic disclaimer must be invoked with particular emphasis on this occasion: ultimate responsibility for the contents of this *Guide* rests entirely with the three named authors, and none of our Working Group colleagues or other advisers and commentators should be assumed to endorse all, or indeed any particular part, of our text.

We welcome further constructive feedback on all four planned *Guides*, information concerning practitioners' experiences of using them, and suggestions for amendments, improvements or other material that could usefully be included. All correspondence should be addressed to:

Royal Statistical Society
Chairman of the Working Group on Statistics and the Law,
12 Errol Street,
London, EC1Y 8LX
or by email to c.g.g.aitken@ed.ac.uk, with the subject heading "Practitioner Guide No.1".

Our intention is to revise and reissue all four *Guides* as a consolidated publication, taking account of further comments and correspondence, towards the end of 2013. The latest date for submitting feedback for this purpose will be 1 September 2013.

Finally, we acknowledge the vital contribution of the Nuffield Foundation[*], without whose enthusiasm and generous financial support this project could never have been brought to fruition.

Colin Aitken, November 2010
Paul Roberts,
Graham Jackson.

[*]The Nuffield Foundation is an endowed charitable trust that aims to improve social well-being in the widest sense. It funds research and innovation in education and social policy and also works to build capacity in education, science and social science research. The Nuffield Foundation has funded this project, but the views expressed are those of the authors and not necessarily those of the Foundation. More information is available at www.nuffieldfoundation.org.

Membership of the Royal Statistical Society's Working Group on Statistics and the Law

Working Group

Colin Aitken, University of Edinburgh, ***Chairman***

Iain Artis, Faculty of Advocates

Graham Cooke, Kings Bench Chambers, Bournemouth

Andrew Garratt, Royal Statistical Society, ***Secretary to the Working Group***

Peter Gill, Centre for Forensic Science, University of Strathclyde

HHJ Anna Guggenheim QC

Graham Jackson, Abertay University and Forensic Science Society

Roberto Puch-Solis, Forensic Science Service

Mike Redmayne, LSE

Paul Roberts, University of Nottingham

Jim Smith, Royal Statistical Society and University of Warwick

Karen Squibb-Williams, Crown Prosecution Service

Peter Stelfox, National Policing Improvement Agency

Corresponding members*:* Bar Council of England and Wales; Crown Office and Procurator Fiscal Service; Law Society of England and Wales; Scottish Police Services Authority

International Advisory Panel

John Buckleton, Institute of Environmental Science and Research, Auckland, NZ

Joe Cecil, Federal Judicial Center, Washington DC

Stephen Fienberg, Carnegie-Mellon University

James Franklin, University of New South Wales, Sydney

Joseph Gastwirth, George Washington University

Jonathan J. Koehler, Arizona State University

Richard Lempert, University of Michigan

Nell Sedransk, National Institute of Statistical Science, Research Triangle Park, NC

Franco Taroni, Institute of Police Science, University of Lausanne

Peter Tillers, Cardozo Law School, New York

1. Probability and Statistics in Forensic Contexts

1.1 ***Probability and Statistics – Defined and Distinguished***

Probability and statistics are overlapping but conceptually quite distinct ideas with their own protocols, applications and associated practices. Before proceeding any further it is vital to define these key terms, and to clarify the relationships between them.

Most of this report is devoted to analysing aspects of probability, more particularly to forensic applications of probabilistic inference and probabilistic reasoning. At root, probability is simply one somewhat specialised facet of logical reasoning. It will facilitate comprehension to begin with more commonplace ideas of statistics and statistical evidence.

1.2 Statistics are concerned with the collection and summary of empirical data. Such data are of many different kinds. They may be counts of relevant events or characteristics, such as the number of people who voted Conservative at the last election, or the number of drivers with points on their licenses, or the number of pet owners who said that their cat preferred a particular brand of tinned cat food. Statistical information is utilised in diverse contexts and with a range of applications. Economic data are presented as statistics by the Consumer Price Index. In the medical context there are statistics on such matters as the efficacy of new drugs or treatments, whilst debates on education policy regularly invoke statistics on examination pass rates and comparative levels of literacy.

Statistics may also relate to measurements of various kinds. Familiar examples in criminal proceedings include analyses of the chemical composition of suspicious substances (like drugs or poisons) and measurements of the elemental composition of glass fragments. Whilst these sorts of forensic statistics are routinely incorporated into evidence adduced in criminal trials, any kind of statistical information could in principle become the subject of a contested issue in criminal litigation. These measurements are sometimes known generically as ‘variables’, as they vary from item to item (e.g. variable chemical content of narcotic tablets, variable elemental composition of glass fragments, etc.).

1.3 Probability is a branch of mathematics which aims to conceptualise uncertainty and render it tractable to decision-making. Hence, the field of probability may be thought of as one significant branch of the broader topic of "reasoning under uncertainty".

Assessments of probability depend on two factors: the event E whose probability is being considered and the information I available to the assessor when the probability of E is being considered. The result of such an assessment is *the probability that E occurs, given that I is known*. All probabilities are conditional on particular information. The event E can be a disputed event in the past (e.g. whether Crippen killed his wife; whether Shakespeare wrote all the plays conventionally attributed to him) or some future eventuality (e.g. that this ticket will win the National Lottery; that certain individuals will die young, or commit a crime).

The best measure of uncertainty is probability, which measures uncertainty on a scale from 0 to 1. In useful symbolic shorthand, x denotes 'some variable of interest' (it could be an event, outcome, characteristic, or whatever), and $p(x)$ represents 'the probability of x'. An event which is certain to happen (or certainly did happen) is conventionally ascribed a probability of one, thus $p(x) = 1$. An event which is impossible – is certain *not* to happen or have happened – has a probability of zero, $p(x) = 0$. These are, respectively, the upper and lower mathematical limits of probability, and values in between one and zero represent the degree of belief or uncertainty associated with a particular designated event or other variable. Alternatively, probability can be expressed as a percentage, measured on a scale from 0% to 100%. The two scales are equivalent. Given a value on one scale there is one and only one corresponding value on the other scale. Multiplication by 100 takes one from the $(0; 1)$ scale to the $(0\%; 100\%)$ scale; division by 100 converts back from the $(0\%; 100\%)$ scale to the $(0; 1)$ scale.

Probability can be "objective" (a logical measure of chance, where everyone would be expected to agree to the value of the relevant probability) or "subjective", in the sense that it measures the strength of a person's belief in a particular proposition. Subjective probabilities as measures of belief are exemplified by probabilities associated with sporting events, such as the probability that Red Rum will win the Grand National or the probability that England will win the football World Cup. Legal proceedings rarely need to address objective probabilities (although they are not entirely without forensic applications).[5] The type of probability that arises in criminal proceedings is overwhelmingly of the subjective variety, and this will be the principal focus of these *Practitioner Guides*.

Whether objective expressions of chance or subjective measures of belief, probabilistic calculations of (un)certainty obey the axiomatic laws of probability, the most simple of which is that the full range of probabilities relating to a particular universe of events, etc. must add up to one. For example, the probability that one of the runners will win the Grand National equals one (or very close to one; there is an exceedingly remote chance that none of the runners will finish the race). In the criminal justice context, the accused is either factually guilty or factually innocent: there is no third option. Hence, p(Guilty, G) + p(Innocent, I) = 1. Applying the ordinary rules of number, this further implies that p(G) = 1-p(I); and p(I) = 1-p(G). Note that we are here specifically considering *factual* guilt and innocence, which should not be confused with the *legal* verdicts pronounced by criminal courts, i.e. "guilty" or "not guilty" (or, in Scotland, "not proven"). Investigating the complex relationship between factual guilt and innocence and criminal trial verdicts is beyond the scope of this *Guide*, but suffice it to say that an accused should not be held legally guilty unless he or she is also factually guilty.

Mathematical probabilities obeying these axioms are powerful intellectual tools with important forensic applications. The most significant of these applications are explored and explained in this series of *Practitioner Guides*.

1.4 The inferential logic of probability runs in precisely the opposite direction to the inferential logic of statistics. Statistics are obtained by employing empirical methods to investigate the world, whereas probability is a form of theoretical knowledge that we can project onto the world of experience and events. Probability posits theoretical generalizations (hypotheses) against which empirical experience may be investigated and assessed.

---

[5] Eggleston (1983: 9) mentions the example of proceedings brought under the Betting and Gaming Act 1960, where the fairness of the odds being offered in particular games of chance was in issue.

Consider an unbiased coin, with an equal probability of producing a 'head' or a 'tail' on each coin-toss. This probability is 1 in 2, which is conventionally written as a fraction (1/2) or decimal, 0.5. Using "p" to denote "probability" as before, we can say that, for an unbiased coin, p(head) = p(tail) = 0.5. Probability theory enables us to calculate the probability of any designated event of interest, such as the probability of obtaining three heads in a row, or the probability of obtaining only one tail in five tosses, or the probability that twenty tosses will produce fourteen heads and six tails, etc.

Statistics, by contrast, summarise observed events from which further conclusions about causal processes might be inferred. Suppose we observe a coin tossed twenty times which produces fourteen heads and six tails. How suggestive is that outcome of a biased coin? Intuitively, the result is hardly astonishing for an unbiased coin. In fact, switching back from statistics to probability, it is possible to calculate that fourteen heads or more would be expected to occur about once in every 17 sequences of tossing a fair coin twenty times, albeit that probability theory predicts that the most likely outcome would be ten heads and ten tails if the coin is unbiased. But what if the coin failed to produce any tails in a hundred, or a thousand, or a hundred thousand tosses? At some point in the unbroken sequence of heads we would be prepared to infer the conclusion that the coin, or something else about the coin-tossing experiment, is biased in favour of heads.

1.5 In summary, probabilistic reasoning is logically ***deductive***. It argues from general assumptions and predicates (such as the hypothesis that "this is a fair coin") to particular outcomes (predicted numbers of heads and tails in a sequence of coin-tosses). Statistical reasoning is ***inductive***. It argues from empirical particulars (an observed sequence of coin-tosses) to generalisations about the empirical world (this coin is fair – or, as the case may be, biased). To reiterate: probability projects itself out onto the empirical world; statistics are derived and extracted from it.

1.6 ***Presenting Statistics***

Statistics that summarise data are often represented graphically, using histograms, bar charts, pie charts, or plotted as curves on graphs. Data comprising reported measurements of some relevant characteristic, such as the refractive index of glass fragments, are also often summarised by a single number, which is used to give a rough indication of the size of the measurements recorded.

1.7 The most familiar of these single number summaries is the ***mean*** or average of the data. For the five data-points (counts, measurements, or whatever) 1, 3, 5, 6, 7, for example, the average or mean is their sum (1+3+5+6+7) divided by the number of data-points, in this case 5. In other words, 22 divided by 5, which equals 4.5.

An alternative single number summary is the ***median***, which is the value dividing an ordered data-set into two equal halves; there are as many numbers with values below the median as above it. In the sequence of numbers 1, 3, 5, 6, 7, the median is 5. For an even number of data points, the median is half-way between the two middle values. Thus for the six numbers 1, 3, 5, 6, 7, 8, the median is 5.5. The mean and median are sometimes known as measures of location or central values.

A third way of summarising data in a single number is the ***mode***. The mode is the value which appears most often in a data-set. One might say that the mode is the most popular number. Thus, for the sequence 3, 3, 3, 5, 9, 9, 10, the mode is 3. However, the median of this sequence is 5, and the mean is 6. This simple illustration contains an important and powerful lesson. Equally valid ways of summarizing the same data-set can produce completely different results. The reason is that they highlight different aspects of the data.

1.8 All of these summaries are estimates of the corresponding characteristics (mean, median or mode) of the population from which the sample was taken. In order to assess the quality of an estimate of a population mean it is necessary to consider the extent of variability in the observations in the sample. Not all observations are the same value (people are different heights, for example). What are known as ***measures of dispersion*** consider the spread of data around a central value. One such measure which is frequently encountered in statistical analysis is the ***standard deviation***. The standard deviation is routinely employed in statistical inference to help quantify the level of confidence in the estimation of a population mean (i.e. the mean value in some population of interest). It is calculated by taking the square root of the division of the sum of squared differences between the data and their mean by the sample size minus one. Large values for the standard deviation are associated with highly variable or imprecise data whereas small values correspond to data with little variability or to precise data. At the limit, if all observations are equal (e.g. every observation is 2), their mean will be equal to each observation (the mean of any sequence of observed 2s is 2). By extrapolation, the differences between each observation and the mean will be zero in every case and the standard deviation will be zero.

To illustrate: consider the sample (set of numbers) 1, 3, 5, 7, 9. The sample size is 5 (there are five members of the sample) and the mean is 5 (1+3+5+7+9 =25; 25/5 = 5). The standard deviation is calculated as the square root of

$$[\{(1-5)^2 + (3-5)^2 + (5-5)^2 + (7-5)^2 + (9-5)^2\}, \text{divided by } 4]$$

which is the square root of

$$(16 + 4 + 0 + 4 + 16)/4 = 40/4 = 10.$$

The square root of 10 is 3.16, which is the standard deviation for this sample set.

By way of contrast, compare the sample (set of numbers) 3, 4, 5, 6, 7. This sample likewise has five members and a mean of 5. However, the standard deviation is much smaller. It is the square root of

$$[\{(3-5)^2 + (4-5)^2 + (5-5)^2 + (6-5)^2 + (7-5)^2\}, \text{divided by } 4]$$

which is the square root of

$$(4 + 1 + 0 + 1 + 4)/4 = 10/4 = 2.5.$$

Thus, the standard deviation is the square root of 2.5 = 1.58. The smaller value for the standard deviation of the second set of numbers reflects reduced variability (illustrated by the reduced range within which the numbers all fall) in comparison to the first sample set.

1.9 *Statistical Method – Sampling and Confidence Levels*

Statistics relate to a designated "population" of relevant events, individuals, characteristics or measurements, etc. Data collection and analysis encompassing every member of a population of interest (an entire set or "***census***") need not involve probabilistic reasoning at all. However, statistics derived from a sample of a larger population can support inferences about the general population only on the basis of probabilistic reasoning.

Suppose that we wish to survey judicial attitudes regarding the reforms of English hearsay law introduced by the Criminal Justice Act 2003. The relevant population is therefore serving judges in England and Wales. Ideally, we might canvass the attitudes of every single judge through a well-designed questionnaire or interview schedule. Having conducted this research project we might discover, say, that overall 73% of judges are in favour of the reforms, but that 80% think they are too complex whilst 14% believe that we would have been better off leaving the old common law unreformed. There is nothing probabilistic about these statistics, because every member of the relevant population was included in the survey (and by redefining "relevant population", probabilistic calculations could still be avoided without conducting a comprehensive census, e.g. "25% of the twenty judges we interviewed thought that…").

More typically, it is impractical to interview every member of a relevant population and insufficiently rigorous simply to interview an arbitrary subset without any consideration of the methodological implications. Resort to some kind of sampling process is consequently almost inevitable.

1.10 Ideally, a good sample is constituted by a "***random sample***" of the ***target population***, i.e. that group of individuals about whom information is sought. In a random sample, every member of the target population has an equal probability of being selected as part of the sample. One must ensure that the population from which the sample is taken (the ***sampled population***) actually is the target population. Imagine an opinion survey for which the target population is all undergraduates at a particular university. Neither a sample of those students arriving at the university library when it opens on a Monday morning, nor a sample of those students propping up the Union bar at 10.00 p.m. on a Saturday night would successfully match the sampled population to the target population. Sometimes a target population may usefully be divided into sections known as ***strata*** defined by relevant characteristics of interest (in a survey to determine whether the population supports a new law concerning sex discrimination one might wish to stratify by gender to ensure that the views of men and women are represented in proportion to their fractions of the population). A ***stratified sample*** contains suitable proportions from each pertinent stratum of the target population.

In practical contexts, including forensic science and criminal litigation, it is often impossible to identify existing random samples or to generate new ones, stratified or otherwise. Instead, resort must be had to ***convenience samples***, that is, samples conveniently to hand. Diamond (2000) calls these data sets "***nonprobability convenience samples***", underlining their acknowledged lack of randomness. Convenience samples might be, for example, "all glass fragments examined in this particular laboratory over the last five years" or "every shoe-mark comparison that I have seen in my career". The methodological robustness of convenience samples and the legitimacy of their forensic applications are perennially debated. Evett and Weir (1998, 45) comment that "every case must be treated according to the circumstances within which it has occurred, and… it is always a matter of judgement…. In the last analysis, the scientist must also convince a court of the reasonableness of his or her inference within the circumstances as they are presented as evidence".

1.11 One form of inference from sample data to a general population is known as ***estimation***. For example, we might seek to estimate the proportion of all judges in favour of the CJA 2003's hearsay reforms by interviewing a sample of judges. The reliability of any such estimate depends on the appropriateness and robustness of the sampling method employed. A carefully constructed random sampling of, say, 10% of all trial judges in every Crown Court is likely to produce more reliable data – i.e., is likely to be more representative of the population as a whole – than taking a straw poll of the first three judges one happens to encounter in the precincts of the Royal Courts of Justice.

1.12 Statisticians employ probabilistic formulae to measure levels of uncertainty associated with particular estimates. Uncertainty is often expressed in terms of "confidence levels". If a sampling procedure produces a particular statistic – e.g. that 75% of judges polled on balance support the CJA 2003's hearsay reforms – how confident can one be that this result is truly representative of the opinions of the entire population of judges? (Recall that the result of our imaginary census of all judges was a 73% approval rating.) Our random sample might have accidentally included judges with more extreme, or more moderate, opinions than their judicial colleagues. Inclusion of these "outliers" would skew our data – but *ex hypothesi* we do not know whether the 75% statistic derived from our sample over- or under-estimates judicial enthusiasm for the CJA 2003, or is in fact truly representative of the opinions of the entire population of trial judges.

By reference to the size of the sample as a proportion of the entire population of interest (in our example, trial judges in England and Wales) and making certain assumptions about variability in responses, it is possible to calculate confidence intervals for the percentage of CJA-supporting judges across the entire population. We know before conducting any survey that the true percentage of judges who favour the CJA 2003's hearsay reforms must logically lie somewhere between 0% and 100%. We could say that we are 100% confident that the true statistic will lie in this range. As the statistical range narrows, our confidence level will diminish. Taking the 75% judicial approval rating as our datum, we can be more confident that the true figure is within the range 75% plus or minus 10% (i.e., the range 65% - 85%) than in the smaller range 75% plus or minus 2% (i.e., the range 73% - 77%).

1.13 Statisticians routinely combine the ***sample mean*** (the mean value for the sample) with the ***sample standard deviation*** to calculate intervals known as ***confidence intervals*** within which the ***population mean*** (the mean value for the entire population) lies with a certain level of confidence. In this context, "confidence" resembles a probability (although its epistemological status is quite different). Confidence levels are usually expressed as a percentage between 0% and 100%. The wider the interval, the greater confidence one has that the stated confidence interval contains the population mean. Confidence intervals are simply a way of representing uncertainty in estimating the population mean.

The only way to be 100% confident that the interval contains the population mean is to make the interval infinitely wide. This is a logical consequence of uncertainty, which can only be (theoretically) eliminated by including every possible value within the interval. Fortunately, we can construct very short intervals with very high degrees of confidence such as 95% or 99%, which are the "gold standard" in social science research and elsewhere. Results falling outside these confidence levels are declared *statistically significant.*

However, confidence intervals and related judgements of statistical significance are not appropriate measures of the value of evidence in criminal proceedings, for several important reasons. First, the selection of a confidence level is subjective and arbitrary. Why 95%? Why not 99% or 99.9%, or for that matter 75% or 70%? Levels of confidence which are conventionally regarded as satisfactory in social science research have no bearing on the level of confidence ideally required for epistemically warranted verdicts in criminal proceedings. Secondly, employing categorical levels of confidence leads to evidence "falling off a cliff" – i.e., it is excluded entirely - if it falls outside the chosen confidence interval, even by a tiny margin. Evidence which may be highly probative within the stated confidence interval is arbitrarily allotted a value of zero if a small change takes it outside that (arbitrarily chosen) confidence interval. Whatever the merits for social science in proceeding in this fashion, it is plainly unsatisfactory for evidence to be allowed to "fall off a cliff" in criminal proceedings, especially when it is recalled that assessments of statistical significance are merely a way of representing variation in data. Consequently, the fact that a particular estimate falls outside one's preferred confidence interval does not necessarily mean that this result is uninteresting or provides an inaccurate measure of real world events which are themselves subject to natural variation.

1.14 ***Statistical Evidence and Inference***

Statistical inference is the science of interpreting data in order to improve our understanding of events in the world, which in turn may contribute to evidence-based public policy-making. For example, statistical inference from meteorological data might help us to understand climate change and to develop more successful strategies for dealing with it. There is an obvious affinity between statistical inference employing probabilistic reasoning (i.e. reasoning employing probabilities) and criminal adjudication, which is also a form of "reasoning under uncertainty" – we do not know whether the accused is guilty or innocent, and the trial is meant to resolve that issue in a publicly acceptable fashion and to translate it into an appropriate legally-sanctioned verdict.

1.15 It is useful, where possible, to be able to measure uncertainty about issues such as guilt or innocence, so that one can compare levels of uncertainty for different events or different pieces of evidence. One might compare, for example, the probability that the accused is guilty, in light of the evidence adduced at trial – conventionally denoted $p(G|E)$ ("the probability of Guilt, given the Evidence"); and $p(I|E)$, the probability of innocence, given the evidence. These are illustrations of the conditionality of specific probabilities to which reference has already been made. The probability of the event of interest – guilt or innocence – is *conditioned on* (assumes) the evidence adduced at trial. Note the use of the vertical bar | to denote conditioning: to the right of the bar is the assumed known (here E, the evidence); to the left of the bar is the uncertain variable for which a probability is being calculated. In relation to fact-finding in criminal proceedings, this will often be G, guilt; or I, innocence. Since it is certain that the accused is either factually guilty or factually innocent (there is no third option), $p(G|E) + p(I|E) = 1$ (meaning that the probability of Guilt, given the Evidence; plus the probability of Innocence, given the Evidence, logically exhausts the range of all eligible possibilities).

Here, uncertainty is a measure of belief in the truth of the matter at issue (e.g. guilt or innocence). The more strongly it is believed that the accused is guilty, the closer that p(G) will approximate to one. In the criminal justice context, the fact-finder's beliefs are ultimately decisive. Note that, whilst the accused is either factually guilty or not (there is no third option), the measure of one's belief in each of the two possible alternatives can be represented by two probabilities taking any value between zero and one. Where there are two exhaustive and mutually exclusive possibilities, the probability of one can always be calculated if the other is known, e.g. p(G) = 1-p(I); and vice versa, p(I) = 1-p(G). Empirical events are *never* absolutely certain, however, so they can only ever *approximate* one (true) or zero (false). This is just another way of saying that reasoning about empirical events is always, irremediably, reasoning under uncertainty.

1.16 Statistical information may be directly relevant to the matters in issue in criminal proceedings, e.g. in assessing levels of risk involved in particular activities such as driving or operating hazardous machinery. If we wish to know whether the accused was reckless or negligent in causing injury to the victim it is pertinent to know the background or "***base rate***" level of risk for that particular activity. If accidents of a particular sort happen all the time, it is so much less likely that the accused was culpably negligent on this occasion. (Base rates are further discussed in section 2(d), below.)

1.17 Statistics are also a useful way of summarising and presenting pertinent information in legal proceedings. For example, large spreadsheets of data may conveniently be summarised in tables or displayed graphically, and this is entirely appropriate provided that such "demonstrative evidence" is properly understood and that its probative value is competently evaluated.

1.18 As well as contributing items of evidence in the form of statistics, statistical methods can also be employed to interpret data and to evaluate evidence. Examples that might well be encountered in criminal litigation include:

- Reliance on statistical evidence of the quantities of drugs on banknotes, to help the fact-finder to assess – relying on an expert's statistical analysis – whether the banknotes are associated with drug dealing (where the quantities of drugs detected are greater than what might be expected for banknotes in general circulation).

- Reliance on statistical evidence comparing the chemical compositions of drugs from two different seizures, to help the fact-finder to assess – relying on an expert's statistical analysis – whether the seized items originated from the same source.

- Reliance on statistical evidence concerning the occurrence of sudden unexplained infant death (in the general population, or amongst families with particular characteristics), to help the fact-finder to assess – relying on an expert's statistical analysis – whether the occurrence of multiple deaths in any one family should be treated as suspicious.

In each of these illustrations (and countless others that might have been given) statistics are being used, not merely as data with evidential significance for resolving disputed facts which could conceivably be adduced in court as expert evidence, but also as a basis for drawing further inferential conclusions the adequacy of which can be assessed by employing statistical methods and probability theory. Insofar as expert testimony incorporates such statistical or probabilistic reasoning, those experts who produce the evidence, those lawyers who adduce and test it, and those judges and fact-finders who evaluate it all need to grasp the rudiments of statistical inference at a level appropriate to their allotted roles in criminal litigation.

1.19 It is useful to distinguish between two types of sample which typically feature in the evaluation of scientific evidence in criminal proceedings. Unfortunately, there is no standard or agreed terminology to express the relevant distinction, which is between (i) samples of known origin and (ii) samples of unknown origin *relative to an issue in the case*. A sample of unknown origin can be described as the ***recovered sample*** or ***the questioned sample***, whereas samples of known origin are often described as the ***control sample*** or ***reference sample***. The issue is not *where* the sample came from, since samples taken from a crime scene (or victim, or abandoned vehicle, etc.) could be either recovered or control samples, depending on the issue being addressed. The objective is normally to link physical traces associated with an offence to the perpetrator, but sometimes this involves working from samples deposited by an unknown donor at the crime scene or on a victim, etc, and sometimes working in the opposite direction, from samples known to be associated with a suspect or victim which can be linked back to the crime scene or suspect, etc.

For example, fragments of glass collected by an investigator from a broken window at the scene of the crime would be a control sample if the question is: does glass found on the suspect's clothing come from the broken window at the scene of the crime? The origin of the fragments is known to be the window. Similarly, a DNA swab taken from a suspect is a control sample as the origin of the profile is known to be the suspect. Suppose in the first case a suspect is found and fragments of glass are recovered from his clothing. These fragments are a recovered sample, since their origin is unknown: it may or may not be the window at the crime scene. Suppose in the second case a DNA profile is obtained from a blood stain at the crime scene. This is also a recovered sample of unknown origin. It may have come from the suspect, innocently or otherwise, or it may have come from another person entirely.

The control/reference sample may have been taken from a crime scene, victim or suspect. Conversely, a recovered/questioned sample might equally derive from any of these sources. Samples are categorised according to the unknown factor the forensic scientist is seeking to investigate, rather than by reference to their physical location and provenance.

1.20 Finally, statistical methods may be utilised to generate new data with forensic applications (although this may be relatively rare in routine forensic science practice). The first task is to define the forensic problem, which initially confronts investigators and is ultimately determined by jurors in contested criminal trials, e.g., have banknotes recovered from the accused been used in drug dealing activity?[6] A determination is then made as to what information is relevant (e.g. to what extent are banknotes in general circulation contaminated with traces of illegal substances?) and this in turn allows the investigator to assess how a reliable sample might be generated in order to produce new data supporting sound inferential conclusions.

[6] Cf. *R* v *Benn and Benn* [2004] EWCA Crim 2100, discussed in §2.22, below.

1.21 We are now beginning to glimpse the power and variety of the potential applications of statistical inference in the administration of criminal justice. It must be stressed, however, that statistical inferences are ultimately only as good as their underlying data, which in turn depends upon (1) the appropriateness of the research design (including sampling methodology) and (2) the integrity of the processes and procedures employed in data collection. Conversely, if data-collection was sloppy and incomplete or samples were poorly chosen, the validity of the inferences drawn from statistical data may be seriously compromised.

1.22 When statistics are being presented and interpreted in forensic contexts (or for that matter, in any other context), there are always two principal dimensions of analysis to be borne in mind:

(i) ***Research methodology and data collection***: Do statistical data faithfully represent and reliably summarise the underlying phenomena of interest? Do they accurately describe relevant features of the empirical world?

(ii) ***The (epistemic) logic of statistical inference***: Do statistical data robustly support the inference(s) which they are said to warrant? Is it appropriate to rely on particular inferential conclusions derived from the data?

2. Basic Concepts of Probabilistic Inference and Evidence

2.1 The first sections of this *Guide* have discussed statistics and statistical evidence in a general way, and introduced some elementary features of probability, including basic notation. In this section and the next we undertake a more systematic and detailed examination of probabilistic reasoning in criminal proceedings.

2.2 The starting point for thinking about information which is statistical or presented in the form of a probability is exactly the same as the starting point for interpreting evidence of *any* kind. The essential issue is: what does the evidence *mean*? The meaning of evidence is a function of the purpose(s) for which it was adduced in the proceedings, which in turn are defined by the issues in the case.

In the context of criminal adjudication, the interpretation of evidence has two principal dimensions. First, the judge must assess whether the evidence is legally admissible. Evidence is admissible if (and only if) it is (i) relevant and (ii) not excluded by an applicable exclusionary rule (such as the hearsay prohibition, rules excluding unfairly prejudicial bad character evidence, or prosecution evidence inconsistent with the demands of a fair trial). Secondly, the fact-finder (jurors or magistrates) must assess the *probative value* of the evidence. This involves determining how the evidence combines with other evidence in the case to support or undermine the prosecution's allegations or the accused's counterclaims. Relevance and probative value are both derived from the logic of inductive inference. Relevant evidence is that which, as a matter of logic and common sense, has some bearing on a fact in issue in the proceedings.[7] The same point can be expressed in terms of probability.[8]

[7] "[T]o be relevant the evidence need merely have some tendency in logic and common sense to advance the proposition in issue": *R* v *A* [2002] 1 AC 45, [2001] UKHL 25, [31] *per* Lord Steyn.

[8] Cf. James Fitzjames Stephen, *A Digest of the Law of Evidence* (Stevens, 12th edn, 1948), Art. 1: "any two facts to which [relevance] is applied are so related to each other that according to the common course of events one either taken by itself or in connection with other facts proves or renders probable the past, present or future existence or non-existence of the other".

Evidence is either relevant or irrelevant, legally speaking. There is no middle ground. Probative value (or the "weight" of the evidence) is the measure of the extent to which relevant evidence contributes towards proving, or disproving, a fact in issue. This is a matter of degree.

2.3 Statistical evidence will be relevant and potentially admissible in English criminal proceedings just insofar as it helps to resolve a disputed fact in issue. Probabilistic reasoning will be useful or even indispensable in criminal proceedings if it is needed to interpret statistical evidence or is otherwise a feature of logical inference and common sense reasoning. In order to interpret and evaluate statistical evidence and to assess the adequacy of any probabilistic inferences which the evidence is said to support, criminal justice professionals need to be familiar with a handful of key concepts that statisticians, forensic scientists, and other expert witnesses use to express probabilities and statistical data. These key concepts include:

(a) (absolute and relative) frequencies;
(b) likelihood of the evidence;
(c) the likelihood ratio;
(d) base rates for general issues (prior probabilities);
(e) posterior probabilities;
(f) Bayes' Theorem; and
(g) independence.

This section will explain and illustrate each of these key concepts in turn. It is perhaps worth reiterating that we are not necessarily advocating any of these approaches to conceptualising evidence and inference in criminal adjudication. It is often possible to arrive at particular inferential conclusions simply by applying inductive logic and "common sense" reasoning without needing to resort to mathematical formulations or consciously-articulated probability calculations. Our aim in describing the intellectual tools examined in this section is to make them more readily accessible to readers who might wish to use them and – no less importantly – to help judges, lawyers and forensic scientists monitor, interpret, evaluate and challenge their use by other professionals in the course of criminal proceedings. Section 3 of this *Guide* extends and reinforces the exposition by identifying common errors ("traps for the unwary") and explaining how to avoid them.

2.4 **(a)** ***Frequencies, relative and absolute***

Frequencies are counts of observed events, characteristics or other phenomena of interest to any inquiry. They answer the question: how often does x occur? Considered in isolation, such counts produce ***absolute frequencies***. However, it is often more useful to ascertain ***relative frequencies***, that is, frequencies relative to a repeated number of observations (e.g. the frequency of rolling a "6" relative to the number of times a six-sided die is rolled). The relative frequency is the number of occurrences of a feature of interest ("rolling a six"; "drawing an ace from a pack of playing cards"; "finding another person with the same DNA profile"; or whatever), divided by the total number of times the process is repeated.

In the forensic context, stated frequencies normally relate to the occurrence of case-specific evidence, whereas frequencies for the occurrence of issues are usually described as ***base rates***. We will have more to say about base rates in due course (§2.16 – §2.18).

2.5 Frequencies can be illustrated by imagining a roulette wheel with thirty-seven slots, numbered 0-36 in the standard pattern. Consider an experiment (or "trial", in the non-legal sense) in which the wheel is spun 1,000 times and the slot on which the ball lands each time is recorded. The number of times on which the ball lands on a particular slot is the absolute frequency for the number corresponding to that slot. Division of the absolute frequency by 1,000 (the number of spins) gives that slot's *relative* frequency. Similar observed frequencies (absolute and relative) can be recorded for each slot. Relative frequencies are often reported as percentages.

For example, in 1,000 spins the ball might be observed to come to rest in the slot numbered one ("slot no.1") 35 times. This is a relative frequency of 3.5% (35 divided by 1,000). In a fair wheel the ball is equally likely to come to rest in any one of the 37 slots, so the expected number of times the ball would come to rest in slot no.1 is one out of every 37 spins, or 1/37 = 2.7%. Statistical methods can then be used to assess the implications (if any) of this evidence of an observed relative frequency of 3.5% in 1,000 spins against a hypothesis that the wheel is fair with an expected relative frequency of 2.7%. One might want to determine, for example, whether the wheel is fair or biased.

2.6 Assessing the adequacy of an inference is never a purely statistical matter in the final analysis, because the adequacy of an inference is relative to its purpose and what is at stake in any particular context in relying on it. A gambler might treat an observed frequency of 3.5% relative to an expected frequency of 2.7% as sufficient reason for putting his money on no.1, but this discrepancy would not warrant a criminal trial jury drawing the inference that the casino owner is guilty of cheating with a biased roulette wheel. In fact, according to probabilistic calculation, one should expect at least 35 slot no.1s in 1,000 spins about once in every 13 sequences of 1,000 spins. The ultimate inferential conclusion that the evidence proves the accused's guilt beyond reasonable doubt or so that the fact-finder is "sure" is never based solely on the probability of any event; not least because fact-finding in criminal adjudication involves normative issues of juridical classification and moral reasoning (Roberts and Zuckerman, 2010: 133-37). However, the inference of guilt beyond reasonable doubt might well be supported – even very strongly supported – by statistical analysis of relevant data and probabilistic reasoning employing absolute or relative frequencies, where the probability of obtaining particular data (evidence) purely by chance is exceedingly small (unlikely, "beyond reasonable doubt"). Imagine, for example, that the accused claimed to have won the National Lottery jackpot five weeks on the trot or that all three of his bigamous wives on whom he had taken out life insurance accidentally drowned in the bath.[9] At some point in the story, "pure coincidence" as an explanation of apparently incriminating circumstances ceases to retain much plausibility – though it is vital to remember that certain kinds of evidence are prone to replicated error (e.g. a string of eyewitnesses might all misidentify an innocent person as the culprit because she does in fact resemble the real offender).

2.7 Spinning a roulette wheel 1,000 times represents a sample subset of the conceptually infinite population of all possible spins of the wheel. The observed frequency of 3.5% is correspondingly an estimate of the true (relative) frequency of the no.1 slot for that wheel, just as a straw poll of voters attempts to sample the voting intentions of the entire electorate. Successive repetitions of 1,000 spins of the wheel (repeat sampling) would almost certainly produce different estimates of the true frequency. This gives rise to some complex issues of sampling, which are addressed in technical Appendix B.

---

[9] *R* v *Smith (George Joseph)* (1916) 11 Cr App R 229, CCA.

What is known as the "***error***" in the estimate is a measure of the differences in the estimates produced by repeat sampling, such as repeated experimental trials each comprising 1,000 spins of a roulette wheel. "Error" here is specialist statistical terminology, not to be confused with the commonplace notion of making mistakes. It is not a "mistake" when the roulette wheel produces three slot no.1s on the trot, though this might not be a very good sample from which to generalise because it is so small. For statisticians, "error" models the natural variation in measurements or counts of empirical phenomena, such as spinning a roulette wheel and recording where the ball lands. Error helps to relate the sample to the population. The error can be determined statistically, and this will give us a measure of the quality or "precision" of the estimate. If the precision of the estimates for every slot were calculated over a series of experimental trials, the strength of the evidence supporting the proposition that the wheel is biased could also be calculated. Note that knowledge of the total number of spins (sample size) is essential in order to assess the precision of an estimate. A trial involving 1,000 spins will produce more precise estimates than a trial involving 100 spins, but less precise estimates than a trial involving 10,000 spins of the wheel. Likewise, an inferential conclusion about the fairness or bias of the wheel will be more reliable if it is based on frequencies with calculated measures of precision for all thirty-seven numbers, and not just for the no.1 slot. All else being equal, more data lead to sounder inferences (although no amount of bad data – e.g. those derived from poorly designed experiments or inappropriate samples – will ever reliably warrant inferential conclusions).

2.8 Relative frequencies may in principle be calculated for any population of items, perhaps conceptually unbounded by size. The items might be each individual spin of a roulette wheel or roll of a die, or types of glass, footwear marks, bloodstain patterns, or DNA profiles – relative to, respectively, all spins of the wheel, all rolls of the die, all types of glass seen in a particular laboratory, all types of footwear seen in a laboratory, all bloodstain patterns observed over a period of time, or all DNA profiles in some defined population.

Relative frequencies always state or assume that there is some reference sample against which the frequency of the event in question may be assessed. A further assumption is that this comparison is illuminating and salient for the task in hand. In the context of criminal proceedings, for example, one would expect that a relative frequency would be capable of supporting an intermediate inference about the strength of evidence bearing on disputed facts, leading to the ultimate inference that the accused is innocent or guilty. Nonetheless, there is ample scope for debating the generation of appropriate and meaningful reference samples, and this has occasionally become a bone of contention in criminal appeals.[10]

2.9 Relative frequencies are routinely included in scientific evidence adduced in criminal proceedings. For example, an expert report might contain statements resembling the following:

- "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."
- "In one survey of men's clothing, bloodstaining of the quantity and in the pattern seen on the defendant's jacket has been found to occur in 1% of jackets inspected in this laboratory."

It is vital for judges, lawyers and forensic scientists to be able to identify and evaluate the assumptions which lie behind these kinds of statistics. The value of the evidence cannot be ascertained unless its meaning is properly understood. For each of these three examples, the size of the reference sample needs to be considered (the actual number of glass samples examined by the relevant laboratory in the last five years; the number of burglaries from which the 2% statistic was derived; the number of jackets in the survey of men's clothing) .

2.10 One might begin by querying the appropriateness of the reference samples. In relation to the first statement, for example: How were the reference glass exhibits selected? Were

[10] *R* v *Benn and Benn* [2004] EWCA Crim 2100 (employing database of banknotes collected from the Bank of England as a reference sample for banknotes in general circulation); *R* v *Dallagher* [2003] 1 Cr App R 12, CA (earprint expert's database comprised a personal collection of about 600 hundred photographs and 300 earprints).

they just those deriving from criminal investigations? Are glass samples from the catchment area of the laboratory relevant to the current investigation? How, if at all, does the frequency of occurrence of glass types examined in a forensic science laboratory help to evaluate a "match" (whatever that means) between glass fragments found on the clothing of a suspect and glass recovered from a crime scene? Evidence reporting a comparison of the elemental compositions of glass fragments from a crime scene and from a suspect's clothing will be conditioned on various factors, such as the precise location from which fragments were recovered (e.g. the surface of an item of clothing).

Likewise in relation to the second statement, one might ask whether the appropriate reference sample should be limited to footwear marks from burglaries. Do burglars prefer particular kinds of footwear? Do footwear sole patterns differ from year to year? They presumably do for men and women, and between age groups. Might a better reference sample be constructed from sales data from leading retailers? ***Production or sales figures*** data may be adduced in evidence in criminal proceedings as proxies for relative frequency of occurrence , e.g. "Between April 2005 and March 2007, 10,000 pairs of shoes of the same sole pattern and size as the defendant's shoes were sold in 10 outlets in the North of England". This example deliberately highlights many of the assumptions that may be embedded in such data. What is the relevance of the specified dates? Why only in the North of England (do people never travel to buy shoes?) And what percentage of the entire market has been cornered by those 10 outlets? (Is it 10 outlets out of 12, or out of fifty?) The adequacy of a reference sample might be challenged on any or all of these grounds. Unless footwear marks taken from burglaries constitute a perfectly representative sample of footwear ownership amongst the general population (which seems rather unlikely and anyway cannot simply be assumed), choosing an alternative reference sample will produce a different relative frequency. So the construction and selection of reference samples could have a major bearing on the way in which statistical evidence is presented and interpreted. Experts in particular fields may be willing and able to advise on the relative strengths and weaknesses of particular reference samples, or may operate with their own assumptions. Ultimately, however, it is for the legal system to determine whether such data adequately support particular inferences for the purposes of criminal adjudication.

2.11 The selection of items submitted to the laboratory for analysis also involves a sampling process amenable to statistical evaluation. For example, a scene of crime officer (SOCO) or forensic scientist will not submit for scientific analysis every fibre, glass fragment, or blood droplet identified at a crime scene, but will instead make selections of samples to be tested. Any sampling process introduces a risk that the sample will be skewed and unrepresentative, but non-randomized samples of this kind require particularly careful scrutiny. (For obvious reasons, such samples are sometime referred to as ***convenience samples***, but this terminology is not employed consistently and forensic scientists may reserve the term for more systematically collected data as opposed to crime-scene samples). As we have seen, the precision of an estimate can be determined statistically, and may be affected by, amongst other things, the size of the sample. If a desirable level of precision is specified in advance, the sample size can be determined accordingly (e.g. forensic chemists can specify how many tablets of a questioned substance need to be submitted for chemical analysis out of the entire consignment of tablets seized by customs officials), though care must always be taken in specifying the precise nature of the inference drawn from any non-random sample.

.

Notice, again, that statistical reasoning is involved at two discrete stages of this evidential process. First, we can ask how representative of the entire population of items is the sample of items submitted for analysis, e.g. how representative is the sample of glass tested at the laboratory of all the glass pieces that were present at the scene from the broken window? If the answer to this question is or may be "not very", any inferences drawn from the evidence produced by the test will be correspondingly weakened, ultimately to vanishing point. Secondly, assuming that the tested items constitute a representative sample of the glass in the broken window, the evidential significance of finding matching fragments of glass on the suspect's clothing must still be assessed. What is the probative value of this finding, for example, if the matching fragments represent a specified percentage of a designated reference sample, such as "7% of reference glass exhibits examined in this laboratory over the last 5 years"?

2.12 Finally, observe that our illustrative statements employ vague concepts such as "pattern" and "quantity" the meaning of which is not self-evident. When is a series of marks a "pattern"? How precise is the measure of "quantity"? Moreover, what is the relationship (if any) between quantity, pattern and activity, e.g. between blood spatter and violent assault? The value of the evidence adduced in any particular trial cannot be determined satisfactorily unless and until these matters are clarified.

2.13 **(b) *Likelihood of the evidence***

Statisticians and forensic scientists sometimes use the phrase "the likelihood of the evidence". This is shorthand for "the likelihood of finding the evidence in the context of the crime scene and the environment of the suspect" (or its contextual equivalents). References to "likelihood" in this context are often synonyms for "probability". For example, the conclusion that "it is very likely that this correlation would be seen if the suspect were guilty" is equivalent to saying that "there is a high probability that this correlation would be seen if the suspect were guilty".

An expert's assessment of the likelihood (or probability) of obtaining particular findings should be based on data relevant to the type of evidence in question. Relevant "data" are of different types. Towards the harder end of the spectrum, experts may be able to draw on extensive surveys, databases or experimentation. At the softer end of the spectrum, the only available relevant data may be the expert's personal experiences and memories of previous casework.[11] The question is not whether "data" can be assigned to one artificial classification or another – "hard" or "soft" – but rather whether the available data constitute an adequate basis for inferring particular inferential conclusions for particular purposes. Irrespective of their quality and status, data enable the expert to assign a likelihood (or probability) for particular findings that is necessarily personal and subjective, even in relation to ostensibly "hard" data.

2.14 For reasons that will become more apparent as we proceed, it is often illuminating and sometimes essential to express the extent to which evidence supports a particular proposition relative to another proposition in terms of the ratio of two likelihoods: (i) the likelihood of the evidence if one proposition is true; and (ii) the likelihood of the evidence if the other proposition is true. In the context of criminal proceedings, one might compare the likelihood of the evidence, given the prosecution's proposition (e.g. that the accused was at the scene of the crime); as against the likelihood of the evidence, given the defence proposition (e.g. that the accused was not at the scene of the crime).

[11] The Court of Appeal recently endorsed expert witnesses' reliance on personal experiences and unpublished studies in *R* v *Weller* [2010] EWCA Crim 1085.

Suppose that a bloody footwear mark taken from the scene of the crime is said to "match" (in some specified sense of what constitutes a "match") the sole of a shoe in the accused's possession. The probability of finding this evidence of a match if that shoe made the mark (which would be the prosecution's proposition) will often come close to 1 (unless, for example, the shoe has been worn for a considerable time after the commission of the offence, in which case the shoe's tread pattern might have been different at the time of the burglary). Crucially, however, the probability of finding the evidence of a match if *another shoe* made the mark (a possible defence proposition) will be more than 0. For a very rare mark, the probability could be miniscule (approaching zero) but in other circumstances it could be closer to 1, e.g. if the vast majority of the burglars in that area wear the same fashionable training shoes.

These two likelihoods (or probabilities) then represent "the likelihood of the evidence if the prosecution's proposition is true" and "the likelihood of the evidence if the defence proposition is true". The relative values of these two likelihoods provide a measure of the meaning and probative value of the evidence. This is usually represented as a ratio known as ***the likelihood ratio***, which is further elucidated in §2.17, below.

In certain scenarios, the likelihood of the evidence if the defence proposition is true is closely related to the frequency of occurrence of the evidence. For example, if the frequency of occurrence of some characteristic, say males with blue eyes, is estimated at 30% (equivalent to a probability of 0.3) for some specified suspect population, then the probability that a particular male suspect would have blue eyes, *on the assumption that this suspect is actually innocent*, is 30% (or 0.3).

2.15 It is not always possible to obtain a good estimate for a population relative frequency based on sample data: relevant datasets may be incomplete or non-existent. In these circumstances, relative frequencies may be replaced by estimates based on an expert's personal experience and knowledge of the type of evidence in question. Here are some examples:

> "This type of glass occurs in about 10% of the glass samples that I have encountered in the course of my work."

Equivalently:

> "In my experience, one in ten of the glass samples that I have analysed at this laboratory have been glass of this type,"

or

> "From my experience of analysing glass samples at this laboratory, the probability of encountering this type of glass is 0.1."

Observe that the first example expresses the expert's conclusion as a percentage, the second as a proportion (or relative frequency), and the third as a probability (or likelihood). In each case, the progression from data (the expert's personal experience) to inferential conclusion (percentage, proportion or probability) is clearly indicated.

2.16 Whenever such percentages, proportions or probabilities are stated, it is imperative to scrutinise the basis on which the subjective assessment has been made. The person asserting the probability or likelihood should be able to justify it by reference to reasonable assumptions. Probabilities representing subjective measures of belief ideally should be formulated in ways which draw attention to their subjectivity, as the following examples demonstrate (with emphasis):

> "*I estimate* the probability (likelihood) of finding this type and size of shoe sole pattern at scenes of burglary in this area as 2% (or 1/50 or 0.02)."

> "If the defendant had not hit the victim, it is *my opinion* that the probability of finding blood-staining of the quantity and in the pattern seen on his jacket is 1% (or 1/100 or 0.01). I base this estimate on data from a survey of men's clothing."

The second example invites follow-up questioning about the nature of the quoted survey, its sampling and other methodological parameters, and its overall adequacy as a reference sample in relation to the issues in the case. There is an apparent implication that if the defendant *had* hit the victim (the prosecution's proposition) there is a probability higher than 1% (and perhaps substantially higher) of finding this pattern of blood-staining. However, this kind of assertion may express little more than a forensic scientist's intuitive inference from experience. Its underlying assumptions must be identified and opened up to critical scrutiny before one can begin to assess the true value of the evidence in resolving disputed facts.

2.17 **(c) *The likelihood ratio***

As previously stated (and as its name transparently implies), the "likelihood ratio" is an expression of the ratio of two relevant likelihoods (or probabilities). Here is one example (with emphasis) that might be encountered in criminal proceedings:

> "The blood-staining on the jacket of the defendant is approximately *ten times more likely* to be seen if the wearer of the jacket had, rather than had not, hit the victim."

Notice that this likelihood ratio expresses the likelihood of *the evidence*, under the two competing propositions, as opposed to the likelihood of *the act* of hitting. It does *not* state that "given the blood-staining on the jacket, it is ten times more likely that the wearer of the jacket had hit, rather than had not hit, the victim", which is an altogether different proposition introducing many more contingencies than the blood-staining evidence *per se* is capable of addressing.

Our initial example states the value of *the evidence* explicitly conditioned on two competing propositions. This exemplifies the kind of statement that a forensic scientist might write in a report or give in oral testimony. The second, reformulated statement addresses the issue of whether the defendant had or had not hit the victim in the context of the evidence of the blood-staining *and* any other relevant evidence in the case. This is the type of question which fact-finders, rather than expert witnesses, should be left to resolve in contested criminal trials.

2.18 Unfortunately, these two types of statement are frequently confused in practice, producing what is popularly (but not very helpfully) known as "the prosecutor's fallacy". This is one of the principal "traps for the unwary", which is fully explained and, hopefully, neutralised in Section 3 of this *Guide*.

2.19 The likelihood ratio can still be calculated when the evidence is in the form of continuous measurements as opposed to discrete events or characteristics. For example, evidence of the refractive index of glass fragments can be derived from a comparison of two sets of measurements: one set from the control/reference sample (e.g. glass from the scene of the crime) and the other set from the recovered/questioned sample (e.g. glass of unknown origin recovered from the suspect's clothing following his arrest). The value of this evidence can be assessed by considering two competing propositions: (i) that all the glass came from the same source; and (ii) that the recovered sample and the control sample did not come from the same source (i.e. the two samples have different sources). The likelihood ratio of the glass evidence is the ratio of: (i) the likelihood of the observed measurements if the two glass samples share a common source; and (ii) the likelihood of the observed measurements if the two glass samples have different sources.

Forensic scientists and other expert witnesses often translate the numerical likelihood ratio into a verbal formulation expressing a measure of strength for a particular proposition. For example, the expert might state that:

> "My findings provide moderate [weak/strong/very strong/etc] support for the theory that the accused, rather than some other person, was the driver of the car used in the robbery."

Alternatively, some experts employ a numerical scale (e.g. a six- or ten-point scale) as a more jury-friendly proxy for the likelihood ratio or as a more intuitive and looser quantification of the probative value of their evidence.[12] In whatever way the likelihood ratio (or other asserted measure of probative value) is translated into evidence, and even if the likelihood is presented in its raw numerical form, it is essential that advocates, judges and fact-finders are able to interpret its true meaning and thereby assess the probative value of the evidence. Experts themselves can and should provide vital assistance by clearly acknowledging their use of a conventional linguistic or numerical scale to express the strength of evidential support, and explaining how it maps onto the likelihood ratio, in their written statements and testimony.

[12] Cf. *R* v *Atkins* [2009] EWCA Crim 1876; *R* v *Shillibier* [2006] EWCA Crim 793; *R* v *Bilal* [2005] EWCA Crim 1555.

2.20 **(d) *Base rates for general variables (prior probabilities)***

Base rates (sometimes also called ***background rates***) are the relative frequencies of variables in a general population before consideration of special circumstances or evidence relating to the case in hand. These do not need to be expressed quantitatively. For example (using fictitious statistics merely for the purposes of illustration):

- "The incidence of death directly attributable to Drug A is 85% of all deaths of abusers of Drug A. "

- "Death attributable to natural hypoglycaemia in elderly non-diabetic patients is extremely rare."

Whereas frequencies relate to specific *evidence*, base rates refer to general *events and other background variables*. This vital distinction further clarifies the respective roles of expert witnesses and fact-finders in criminal proceedings. Base rates for general variables are independent of case-specific information, to which they form the backdrop. Thus, base rates may well be introduced by a competent expert before another expert presents, and takes into account, the results of their own examinations. Base rates can also be used to assign ***prior probabilities*** for those events. The term "prior" encapsulates the fact that such probabilities are developed *prior* to any evidence specific to the instant case.

The first question for the expert (or for the first expert) in each of our examples would be 'what is the base rate for the event or characteristic in question (general prevalence among Drug A abusers of death directly attributable to Drug A; incidence of death caused by natural hypoglycaemia in elderly non-diabetic patients)'? The second question for the expert (or the question for the second expert) is 'what is the probability of their findings, taking account of the prosecution's and defence's competing propositions'? In the first example, the expert would need to consider data on (i) levels of Drug A found in drug-abusers who had died as a consequences of ingesting Drug A; and (ii) levels of Drug A found in drug-abusers who had not died as a direct result of ingesting Drug A. In our second example, the expert would derive a likelihood ratio from data on (i) the levels of insulin found in elderly non-diabetic patients who had died through natural hypoglycaemia; and (ii) data relating to deaths in similar patients resulting from induced hypoglycaemia.

Finally, the question for the fact-finder (*not* the expert) in each case is 'what is the probability, in light of the evidence adduced at trial, that the accused is guilty (that the death was directly attributable to the drug; that the elderly deceased was injected with insulin, etc.)'?

2.21 Base rates can have significant implications for inferential conclusions. Imagine a medical diagnostic test with a high probability of a positive result if the patient has the disease (this measure is known as ***sensitivity***) and a high probability of a negative result if the patient does not have the disease (known as ***specificity***). This hypothetical diagnostic test, then, is both very sensitive to the presence of the disease and very specific to it. Suppose that a particular patient is diagnosed with the disease. What is the probability that the patient actually has the disease? The fact that the diagnostic test is both very sensitive and very specific does not, as might be thought, guarantee that a positive diagnosis is very likely to be correct. This is a function of base-rates. Imagine that nobody in the region has the disease (the base-rate is zero). No matter how sensitive and specific the diagnostic test is – perhaps it only errs one time in a million – on this assumption every single positive diagnosis will be wrong. The probability of a correct diagnosis when the base rate is zero is zero, irrespective of the diagnostic power of the test.

2.22 Base rates that are derived from samples (as distinct from those derived from a census) invite methodological questions paralleling those concerns previously identified in relation to calculating the relative frequencies of evidence. Base rates will be a poor base-line for any inferential purpose if data collection was poorly executed or the sampling procedure was methodologically flawed. Even if base rates supply methodologically robust information for some purposes, they will not necessarily serve to illuminate the matters specifically in issue in criminal proceedings. As in relation to frequencies of evidence, one must carefully scrutinise the inferential link, if any, between background base rates and the issues requiring proof in the current trial.

A different sample drawn from the same population will, almost certainly, give a different answer for a relative frequency. This does not mean that either or both of these frequencies is "wrong". Rather, both frequencies are (different) estimates of a true unknown rate. In *Benn and Benn*,[13] a case concerned with base rates for trace quantities of drugs on banknotes in general circulation, the Court of Appeal remarked that, "the question of the validity of a database depends upon the purpose which is to be served".[14] Deficiencies in the database were not considered fatal to the safety of the convictions where "the comparison made between the notes in the appellants' possession and the database was merely part of the prosecution case showing a connection between the appellants and the cocaine". Whilst the value of the statistical evidence was thus marginalised as merely part of the general background to the prosecution's case, the Court did not really consider that sampling deficiencies arguably robbed this evidence of any discernable meaning or probative value.

As we saw in 1.13, the reliability of an estimate can be determined by specifying, at a stated level of confidence (e.g. 95% or 99%), an interval within which the true rate is thought to fall. The narrower the interval for a given confidence level, the more reliable the estimate.

2.23 **(e) *Posterior probabilities***

All probabilities are predicated (or "conditioned") on specified assumptions. This is merely another way of expressing the inherent conditionality of probability as a species of reasoning under uncertainty. Thus, for example, one might calculate the probability that the accused is guilty, given the evidence that has been presented in the trial – in mathematical notation, $p(G|E)$. Whereas base rates for general variables inform prior probabilities, conditional probabilities conditioned on case-specific events or evidence can be described as ***posterior probabilities*** – such as the probability that the accused is guilty after (posterior to) having heard all the evidence. The ultimate posterior probability, of guilt or innocence and their corresponding legal verdicts, is always a question for the fact-finder in English and Scottish criminal proceedings.

2.24 Expert witnesses must not trespass on the province of the jury by commenting directly on the accused's guilt or innocence, and should generally confine their testimony to

---

[13] *R* v *Benn and Benn* [2004] EWCA Crim 2100.

[14] ibid. [44].

presenting the likelihood of their evidence under competing propositions. However, experts are not absolutely precluded from stating posterior probabilities relating to intermediate facts proving or constituting the offence, if invited to do so by the court and providing that such statements are appropriately qualified and contextualised. The court must understand, and be prepared to accept, the suppositions on which statements such as the following are predicated:

- "In my opinion, it is highly likely that the defendant kicked the victim."

- "I believe there is a 99% chance (probability of 0.99) that the defendant handled explosives."

- "In my opinion, the accused is very likely to have been the author of the ransom note."

All these statements relate specifically to evidential facts and only indirectly to the ultimate issue of guilt or innocence. It may be helpful, in appropriate cases, for expert witnesses to express their conclusions in this form (also note that our examples commendably flag up the subjective nature of the inference as the expert's "opinion", "belief", etc.). However, it is vital to appreciate that posterior probabilities relate to disputed facts rather than to information adduced in evidence, and the two must never be confused. Experts normally testify to ***relative frequencies*** (to inform likelihoods of the occurrence of evidence), or occasionally to ***base rates*** (***prior probabilities***), rather than to the truth or falsity of contested issues in the trial (***posterior probabilities***). Where experts depart from the norm by testifying directly to posterior probabilities, they should do so deliberately and advisedly, not merely through confusion. Insofar as experts do testify to posterior probabilities, they must spell out and justify the conditioning assumptions and prior probabilities supposedly warranting them.

2.25 **(f)** ***Bayes' Theorem***

Bayes' Theorem is a mathematical formula that can be applied to update probabilities of issues in the light of new evidence. One begins with a prior probability of an issue and some pertinent item of evidence. Bayes' Theorem calculates a posterior probability for the issue, conditioned on the combined value of the prior probability *and* the likelihood ratio for the evidence. This posterior probability can then be treated as a new prior probability to which a further additional piece of evidence can be added, and a new posterior probability calculated (now taking account of the original prior probability and the likelihood ratios for both pieces of evidence). The process can be repeated over and over, finally resulting in a posterior probability conditioned on the entire corpus of evidence in the case.

Fact-finding in criminal adjudication is, generally speaking, accomplished by ordinary common sense reasoning rather than through the application of mathematical formulae, as the Court of Appeal emphatically reiterated in *Adams*.[15] It should be borne in mind, however, that although most evidence adduced in criminal proceedings does not come with a pre-assigned quantified numerical value attached (e.g. what is the probability that an eyewitness identification is accurate? Or the probability that a confession is true?), much forensic science evidence (including DNA profiling) *is* predicated on quantified probabilities and is consequently directly amenable to Bayesian calculations. Moreover, even unquantified evidence can be assigned a subjective probability in Bayesian reasoning. Bayes Theorem is a codification of the reasoning that should be applied in the assessment of evidence. It is a statement of logic. Its application ensures evidence is assessed rationally.

2.26 Bayes' Theorem is best illustrated through a simple artificial example. Consider a population of interest comprising 1,000,001 people. One person has committed a burglary, the other million are innocent. Suppose that by chance 1% of the innocent people (1,000,000/100 = 10,000) have carpet fibres on their clothing matching the carpet at the burgled premises. Assume that the burglar's clothing also picked up these fibres during the burglary. These distributions are summarised in Table 2.1:

[15] *R* v *Adams (No 2)* [1998] 1 Cr App R 377, CA.

Table 2.1: **Numbers of innocent and guilty people on whom fibres are present and absent**

Fibres	Guilty	Innocent	Total
Present	1	10,000	10,001
Absent	0	990,000	990,000
Total	1	1,000,000	1,000,001

We can read off from the final right-hand column of the 'Present' row that the fibres were found on 10,001 individuals – 10,000 of whom are innocent and one of whom is the guilty burglar.

From these data we can construct ***prior probabilities*** for guilt and innocence, before the evidence of the fibres is considered. The prior probability of guilt is 1/1,000,001 – one person out of 1,000,001 is guilty. In other words, the probability that a person selected at random from the population would be guilty is 1/1,000,001. Complementarily, the probability that a person selected at random from the population is innocent is 1,000,000/ 1,000,001.

The ***posterior probabilities*** for guilt and innocence can be obtained from the row labelled "Present" in which there are 10,001 people of whom 1 is guilty. Thus, after the evidence of the fibres is considered, the posterior probability of guilt is 1/10,001 – one person out of 10,001 is guilty. In other words, the probability a person selected at random from the population on whom relevant fibres are found would be guilty is 1/10,001. Complementarily, the probability that a person selected at random from the population on whom relevant fibres are found is innocent is 10,000/10,001.

The ***likelihood ratio*** is the ratio of the probability for the presence of the relevant fibres amongst the guilty (the proportion of people in Table 2.1's Guilty column for whom the fibres are present) to the probability for the presence of the relevant fibres amongst the innocent (the proportion of people in Table 2.1's Innocent column for whom the fibres are present). In this simple example, the probability for the presence of the relevant fibres amongst the guilty is one divided by one, i.e. 1. The probability for the presence of the relevant fibres amongst the innocent is 10,000 divided by 1,000,000, or 1/100. The ratio of these probabilities is 1 divided by 1/100 which is 100. This may be summarised in words as "the evidence of the presence of relevant fibres is one hundred times more likely if the person is guilty than if the person is innocent".

2.27 These probabilities can also be expressed, equivalently, in terms of ***odds ratios***. The prior odds a person selected at random from the population is guilty are given by the ratio of the two prior probabilities for guilt and innocence, namely 1/1,000,001 divided by 1,000,000/1,000,001 or 1 to 1,000,000. This equates to saying that, "the odds are one million to one against guilt for a person selected at random from the population"; or that "the odds are one million to one in favour of innocence for a person selected at random from the population".

The posterior odds that a person selected at random from the population on whom relevant fibres are found is guilty are given by the ratio of the two posterior probabilities, namely 1/10,001 divided by 10,000/10,001 or 1 to 10,000. This equates to saying that, "the odds are ten thousand to one against guilt for a person selected at random from the population on whom relevant fibres are found"; or that "the odds are ten thousand to one in favour of innocence for a person selected at random from the population on whom relevant fibres are found".

2.28 Bayes' Theorem links the prior odds, the posterior odds, and the likelihood ratio in the following way:

$$\textit{posterior odds} = \textit{likelihood ratio} \times \textit{prior odds.}$$

That is to say, the posterior odds are calculated by multiplying together the likelihood ratio and the prior odds (or again, the posterior odds are the product of the likelihood ratio and the prior odds).

In our example, the prior odds are one in a million, the posterior odds are one in ten thousand and the likelihood ratio is one hundred. This is a verification of Bayes' Theorem, since one in ten thousand is 100 times one in a million. Of course, it is not necessary to apply the sledgehammer of Bayes' Theorem to crack this simple example, the results of which could be obtained more or less directly by common sense mathematical calculation.

Bayes' Theorem comes into its own, and may have significant forensic applications, when the calculations are more complex and the issues to be addressed may not be so self-evident.

2.29 Bayes' Theorem can be expressed more formally and in a way which applies directly to criminal proceedings, as follows:

The ***posterior odds*** in favour of the prosecution proposition are equal to the product of:

(i) the ratio of the probability of the evidence if the prosecution's proposition is true, to the probability of the evidence if the defence proposition is true (i.e. ***the likelihood ratio***); and

(ii) the ***prior odds*** in favour of the prosecution proposition.

Referring back to Table 2.1, the evidence is the presence of fibres on the clothing of a suspect (recovered sample) that are of the same type and colour as carpet fibres at the crime scene (control sample). The prosecution proposition is that the suspect is guilty of the crime. The defence proposition is that the suspect is innocent. The likelihood of the evidence given (conditioned on) the truth of the prosecution's proposition is 1, or $p(E|G) = 1$. The likelihood of the evidence given (conditioned on) the truth of the defence proposition is 10,000/1,000,000 = 1/100, or $p(E|I) = 1/100$. The probability of guilt given the evidence – $p(G|E)$ – is 1/10,001. The probability of innocence given the evidence – $p(I|E)$ – is 10,000/10,001.

Notice that the first pair of quantities is conditioned *on the assumption of guilt* or *of innocence* (as the case may be), whereas the second pair of quantities is conditioned *on the evidence*. Moving from the first pair to the second pair of quantities involves ***transposing the conditional***. It can be see that "E", representing the evidence, occupies the position to the left of the conditioning bar in the first pair of quantities, whereas in the second pair its position has shifted (been transposed) to the right of the bar. Bayes' Theorem can be described as a logical and legitimate procedure for transposing the conditional. Illegitimate transposition of the conditional is (for better or worse) widely known as "the prosecutor's fallacy", which is explained and debunked in Section 3 of this *Guide*.

2.30 A second illustration demonstrates the power of Bayes' Theorem as a formula for updating conditional probabilities and should help to clarify its current and potential forensic applications.

Suppose that an accident is caused by an unidentified bus. A total of 1,000 buses are in service in the vicinity. Blue Bus Company owns 90% of these 1,000 buses and Red Bus Company owns the remaining 10%. An eyewitness testifies that the bus that caused the accident was Red. However, a psychologist gives uncontradicted expert testimony that eyewitness identifications of this type are accurate only about 80% of the time. That is to say, an eyewitness will report seeing a Red (or Blue) bus when the bus was truly Red (or Blue) 80% of the time. Conversely, the eyewitness will report that the bus was Red when it was Blue (or Blue when it was Red) 20% of the time.

The entire population of interest comprises 900 Blue buses (90% of 1,000) and 100 Red Buses (10% of 1,000). If the accident was in fact caused by a Blue bus, the eyewitness would accurately report 720 Blue buses (80% of 900) and misidentify the other 180 (20% of 900) as Red. If the accident was in fact caused by a Red bus, the eyewitness would accurately report 80 Red buses (80% of 100) and misidentify the other 20 (20% of 100) as Blue. On this scenario, a Red bus is four times more likely to be reported as Red than Blue. However, *a priori* there are nine times as many Blue buses as Red buses operating in the area. These results are summarised in Table 2.2.

Table 2.2: Numbers of Red and Blue buses, as reported and in fact

	Actually Blue	Actually Red	Total
Reported Blue	720	20	740
Reported Red	180	80	260
Total	900	100	1000

2.31 Bayes' Theorem states that the posterior odds are equal to the likelihood ratio multiplied by the prior odds.

The prior odds are 9:1 (or, simply, 9) in favour of a Blue bus having caused the accident, or p(Blue) = 0.9. Complementarily, p(Red) = 0.1. The prior odds in favour of a Red bus are the reciprocal of the odds in favour of a Blue bus and are hence 1:9 (or 1/9).

We are told that the eyewitness testifies that the bus involved in the accident was Red. The likelihood ratio is the probability that a bus is reported as Red given that it is Red (80/100) divided by the probability that it is reported as Red given that it is Blue (180/900), which equals 4. The posterior odds of the bus being Red when reported Red are the product of the prior odds and the likelihood ratio, 1/9 multiplied by 4 which equals 4/9. The corresponding probability of a bus being Red given it is reported as Red is then 4/13 and the probability of a bus being Blue, given it is reported as Red is the complement of this, namely 9/13 (Check that the ratio of these two probabilities is 4/9 (or 4:9), the odds.)

It seems counterintuitive that the evidence should favour the bus being Blue when the eyewitness testified Red: but Table 2.2 and Bayes' Theorem both corroborate that conclusion. It is obvious at the outset that – all else being equal – a Blue bus was much (nine times) more likely to be involved in the accident than a Red bus. The eyewitness testimony decreases these prior odds to posterior odds of 9:4. Nonetheless, given the eyewitness's stipulated error rate (20%), when the eyewitness testifies Red this actually favours Blue by a ratio of 180:80, or 9:4 – as can be read off from the "Reported Red" row of Table 2.2. Bayes' Theorem powerfully confirms this counter-intuitive result. The likelihood ratio of 4 reduces the odds in favour of Blue from 9:1 to 9:4. In other words, the eyewitness evidence supports the proposition that the bus is Red, but not with sufficient probative force to make it *more likely than not* that the bus is Red, all things considered. This would require a probability greater than 0.5, or (equivalently) odds greater than 1:1 – "a fifty-fifty chance", as we might say. (Note, however, that this conclusion is alarming for real-world litigation only on the supposition that eyewitnesses really do confuse red and blue 20% of the time, and – to our knowledge – there is no empirical evidence warranting that assumption.)

The purpose of the example is two-fold. First, it provides a numerical verification of Bayes Theorem. Second, it shows how consideration of uncertainty about the accuracy of an eyewitness may be included in the evaluation of the evidence of the eyewitness. One can model the effect of various levels of uncertainty on the value of the evidence of the colour of the bus that was involved in the accident.

2.32 **(g)** ***Independence***

The concept of independence is central to both legal proof and mathematical probability. In law, two or more independent items of evidence may be mutually corroborative. This is first and foremost a logical rule of inference which the law of criminal procedure sometimes elevates into a formal legal requirement (most formal corroboration requirements have been abolished in England and Wales, but Scottish law still retains a general demand for corroboration in serious criminal cases). The logic of corroboration through independent evidence extrapolated to probability theory by ***the product rule for independent events***, which states that, if two events are independent, the probability of both of them occurring together (known as their *conjunction*) can be calculated by multiplying together the probability of the first event and the probability of the second event. These propositions are best demonstrated through simple illustrations using coin-tossing and playing cards.

2.33 Two events are independent in the probabilistic sense if the occurrence of one has no bearing on the probability of the occurrence of the other. Successive outcomes of the tosses of a coin or of tosses of several different coins are independent. Consider two fair coins, which when tossed are (by definition, as "fair" coins) equally likely to produce a head or a tail. The occurrence of a head when the first coin is tossed has no effect on the probability of a head when the second coin is tossed. On the toss of the first coin, the probability of a head is equal to the probability of a tail, which equals ½ or 0.5. These probabilities remain the same for the toss of the second coin and on subsequent tosses of these or of other fair coins. Independence holds no matter how many times the process is repeated.

Consider one fair coin which is tossed twice. The probability of two heads in two tosses of the coin is (utilising the product rule) ½ x ½ = ¼, or p(two heads) = 0.25. The probability of two tails is exactly the same. However, the probability of one head and one tail is ½, or p(head and tail) = 0.5. This is because there are two ways in which the outcome of the two tosses of the coin can be a head and a tail: head followed by tail, or tail followed by head. The probability of each of these two sequences is ¼, and the probability of either one or the other (known as their ***disjunction***) is calculated by adding (*not* multiplying) the probability of each, as we say, *exclusive* event. In other words, ¼ + ¼ = ½. The events (head followed by tail) and (tail followed by head) are known as exclusive events since one occurs to the logical exclusion of the other.

2.34 Now consider a slightly more complicated example. A normal pack of playing cards contains 52 cards in four suits, spades (♠), hearts (♥), diamonds (♦) and clubs (♣) with thirteen cards in each suit. The pack is well-shuffled. A card is picked from the pack at random, i.e. in such a way that each card is equally likely to be selected. Suppose that the ace of spades (*A*♠) is the card drawn at random from the pack. This card is replaced, the pack is well-shuffled and then a second card is drawn, again at random. This process of selection is described as ***selection with replacement***. These successive draws of cards are independent events. Replacing the first card drawn and then shuffling the pack ensures that the outcome of the first draw has no effect on the outcome of the second draw. In other words, the outcomes of the two draws are independent. The probability that the card drawn at the second draw is also the *A*♠ is the same as the probability that the first card was the *A*♠, 1/52. Given that these are (as we have stipulated) independent events, the product rule applies, so that the probability of drawing the *A*♠ twice in succession is 1/52 x 1/52 = 1/2,704.

The same type of calculation can be extended to groups of cards. For example, the probability that a card picked at random from a pack is a ♠ is 13/52 = ¼, or p(spade) = 0.25. There are 13 ♠ in the pack, and each is equally likely to be selected.

2.35 If two or more events are not independent, then they are ***dependent***. There is also a product rule for calculating the probability of the conjunction of dependent events. Consider again the selection of two cards at random from a normal pack, one after the other. This time, the first card selected is *not* returned to the pack after it has been viewed, so that the second card is drawn from a reduced pack of 51 cards. This type of selection process is called ***selection without replacement***.

What is the probability of selecting two aces without replacement? It is the product of the following two probabilities:

(i) the probability that the first card selected is an ace, which is 4/52 = 1/13; *and*

(ii) the probability that the second card selected is also an ace, which is 3/51 = 1/17 (since there are now only 51 cards remaining in the pack, of which 3 are aces).

Thus, the combined probability, p(drawing two aces without replacement) is 1/13 x 1/17 = 1/221.

This result can also be derived and demonstrated by direct enumeration. The order in which the cards are drawn is significant. There are twelve ways of drawing two Aces, viz (♠♥), (♥♠), (♠♦), (♦♠), (♠♣), (♣♠), (♥♦), (♦♥), (♥♣) (♣♥), (♦♣) and (♣♦). There are 52 ways of choosing the first card without replacement and 51 ways of choosing the second card from the reduced deck. There are therefore 52 x 51 = 2,652 equally likely ways of choosing two cards from a pack of 52 cards. Of these 2,652 ways, twelve give two aces. Thus the probability of drawing two aces equals 12/2,652 = 1/221.

2.36 In this example, the probability of the second event (drawing an Ace) is dependent on the first event (drawing an Ace). The probability of drawing a second Ace (i.e., assuming an Ace was drawn on the first draw) is 1/17. This is less than the probability (1/13) of drawing an Ace on the first draw. However, two events may also be associated in such a way as to *increase* the probability of the second event relative to the probability of the first event. This somewhat counterintuitive result is illustrated in Appendix B.

2.37 The probabilistic foundations of games of chance have real-world analogues in criminal litigation. It is therefore vital for criminal practitioners to grasp the fundamentals of probabilistic thinking, and these fundamentals include the concept of independence. The nature of the dependency in examples involving packs of cards or tosses of coins is readily identifiable. In real life the dependencies are typically more difficult to ascertain. Yet as Section 3 will elucidate, it is a serious error to apply the simple product rule to events that are not, or may not be, independent. As a general rule of thumb, independence should be verified and demonstrated and not merely *assumed* by default.

3. Interpreting Probabilistic Evidence – Anticipating Traps for the Unwary

3.1 Reasoning errors in criminal adjudication are by no means confined to information concerned with probabilities. However, probability, statistical evidence, and inferential reasoning associated with them do seem to be especially prone to recurrent errors and misinterpretation. Statistical and probabilistic evidence are typically adduced in court through the medium of a scientific report or expert witness testimony adduced at trial. There is consequently considerable overlap between an examination of probabilistic evidence and reasoning in criminal proceedings and the general topic of expert evidence, as previous sections have already intimated.

3.2 This section begins by drawing attention to some fundamental principles for correctly interpreting reports or testimony provided by forensic scientists and other expert witnesses. We will emphasise, in particular:

(a) the importance of correctly identifying the level of the propositions addressed by the evidence, in order to interpret its real bearing (if any) on the issues in the case; and

(b) the nuanced language used by scientists to express their inferential conclusions, which requires a certain amount of "unpacking" in order to decode its true meaning.

Thereafter, the following analytically distinct (though in practice, often compounded) reasoning errors will be examined and elucidated:

(c) illegitimately transposing the conditional ("the prosecutor's fallacy");

(d) source probability error;

(e) underestimating the value of probabilistic evidence;

(f) probability ("another match") error;

(g) numerical conversion error;

(h) false positive fallacy;

(i) fallacious inferences of uniqueness; and

(j) unwarranted assumptions of independence.

Learning about these reasoning errors as an abstract intellectual exercise is not the same as successfully avoiding them in practice. Their twisted logic can seem enormously seductive and they are frequently perpetrated by professionals who ought to know better, especially in pressured situations such as giving evidence in criminal trials. This is all the more reason for lawyers and judges, as well as forensic scientists and other expert witnesses, to study the recurrent errors in probabilistic reasoning examined in this section. Forewarned is forearmed. Knowing what to look out for, coupled with eternal vigilance, is the best way to guard against falling into traps for the unwary.

3.3 **(a)** ***Relating the evidence to the issue: what question does the expert's evidence purport to answer?***

Expert evidence (or indeed, any other evidence adduced in criminal proceedings) might be conceptualised as offering an answer to a question. The ultimate question in criminal adjudication is always: is the accused guilty or innocent of the offence(s) charged? Of course, in deference to the presumption of innocence the ultimate question in English and Scottish criminal proceedings is not framed in this way. Instead, we ask: has the prosecution proved the accused's guilt beyond reasonable doubt (or so that the fact-finder is "sure" of the accused's guilt)?

Expert evidence does not answer the ultimate question directly; this is a matter solely within the province of the fact-finder. Instead, expert evidence addresses intermediate evidential facts with a bearing on the ultimate issue. For example, an expert might testify that glass found on the accused's clothing resembles (or "matches") glass from the scene of the crime; or that the accused's fingerprints are similar to (or "match")[16] those on the window of the burgled house; or that the type of firearm discharge residue (FDR) evidence found on the victim of a shooting supports the proposition that the accused's gun fired the

[16] The notion of a forensic science expert "declaring a match", though familiar, is problematic. In the first place, the criteria for declaring "a match" may be contested amongst practitioners, or may be eminently contestable even where most or all competent practitioners agree on conventional criteria for determining what constitutes a match. More fundamentally, if all trace evidence ultimately rests on probabilistic calculations, experts perpetrate source probability error (discussed in (d), below) whenever they conclusively assert "a match".

fatal shot. It is then a matter for the fact-finder to determine whether this evidence, taken together with all the other evidence in the case, is sufficient to warrant a finding of guilt.

When one grasps that evidence (including expert evidence) is adduced by the prosecution or defence to answer a particular question, it follows that the meaning and value of that evidence cannot be determined without first identifying the original question. One cannot assess whether evidence is successful in proving a matter in issue until one knows what the issue is and how the evidence relates to it. This observation might sound banal; but it is not. In fact, nearly all of the reasoning errors described in this section are either variations on, or are at least exacerbated by, an elementary failure to identify, with sufficient care and particularity, the question which the evidence is capable of answering.

3.4 A useful starting point in evaluating expert evidence is to identify the ***level of proposition*** (or type of answer) which the evidence addresses. Four different levels of proposition can usefully be distinguished:

(i) source level propositions;
(ii) sub-source level propositions;
(iii) activity level propositions; and
(iv) offence level propositions.

Each of these levels of proposition is regularly encountered in criminal litigation.

3.5 The following are examples of pairs of complementary ***source level propositions***:

- "The defendant is the source of the semen at the crime scene."/
  "The defendant is *not* the source of the semen at the crime scene."

- "The defendant's sweater is the source of the fibres at the crime scene."/
  "The defendant's sweater is *not* the source of the fibres at the crime scene."

- "The damaged window frame is the source of the paint fragments recovered from the defendant's clothing."/
  "The damaged window frame is *not* the source of the paint fragments recovered from the defendant's clothing."

The value of evidence adduced in support of source level propositions is usually related to the relative frequency of the characteristic of interest. Suppose this frequency is one in a thousand (1/1,000). As a first approximation, the value of evidence can be expressed as the reciprocal ("one over") of that relative frequency, e.g. one divided by 1/1,000 or 1,000. For each of our three pairs of example source level propositions, there must be some reference sample (e.g. a database of DNA profiles; records of fibres recovered from crime scenes; or previous analyses of paint fragments found on clothing examined at the laboratory) allowing the expert to calculate the probability of the evidence if it came from an alternative source consistent with the accused's innocence. Notice that source level propositions do not say anything about how the evidence came to be at the scene or on the defendant's clothing, nor do they take into account such variables as the quantity, position or distribution of the recovered material. Source level propositions are limited to addressing whether or not a piece of evidence came from a particular source. Assessment of evidence under source level propositions requires little in the way of circumstantial information.

3.6 Certain forensic science techniques, notably DNA profiling, have become so sensitive that it may be desirable to formulate expert evidence with greater circumspection and precision in terms of ***sub-source level propositions***, such as the following:

- "The DNA recovered from the crime sample came from Mr Smith."
- "The DNA recovered from the crime sample did not come from Mr Smith;" or "The DNA recovered from the crime sample came from some other person."

Sub-source level propositions introduce a greater degree of caution by taking the inferential process, as it were, one stage further back. The expert does not make any direct assertion about the type of biological material from which the DNA was ostensibly extracted (e.g., the semen or blood recovered from the crime scene). Rather, the evidence is restricted to the sub-source or cellular level – leaving open the possibility that the material from which the DNA has been extracted may not be the assumed, asserted or most obvious source. For example, biological samples recovered from the crime scene might contain mixtures of different types of cellular material – saliva, skin cells, secretions, etc. – contributed by several human donors. In these situations, it will be very unlikely that the scientist is able to attribute the DNA to any one type of cellular material.

3.7 Running in the opposite direction, activity level propositions are more coarse-grained and potentially provide more probative evidence than source level propositions. The following are examples of ***activity level propositions***:

- "The defendant had intercourse with the victim."
- "The defendant walked on the carpet in the burgled house."
- "The defendant smashed the window."

The expert is now addressing the issue of whether or not the accused actually did something (had intercourse; walked on a carpet; smashed a window, etc.), not merely whether or not physical evidence might have come from a specified source or sub-source. This is unavoidably controversial territory. In order to arrive at the value of the evidence assuming an activity level proposition, the expert needs to factor into their analysis much more than merely relative frequencies. For example, it may be necessary to consider the physics of transfer and persistence of physical evidence, with associated subjective probabilities. It is also necessary to take into account any innocent explanations offered by the accused for the existence of apparently incriminating evidence. For example, an accused may say that his clothing had been sprayed by the victim's blood when he, an innocent passerby, attempted to render first aid to the dying victim. The scientist's role in this situation is to assess the likelihood of obtaining the pattern and distribution of blood-staining that had been observed on the clothing if the accused's suggestion were, or might have been, true.

Crucially, in terms of the balance and usefulness of scientific findings, consideration of activity level propositions provides an assessment of the probative value of the absence of material ("missing evidence"); something that cannot be assessed if source (or sub-source) propositions are considered.

3.8 ***Offence level propositions*** are the most coarse-grained and probatively consequential of all the types of statement that might be encountered in expert witnesses' reports or testimony. They take the following form:

- "The defendant raped the complainant."
- "The defendant burgled the house."
- "The defendant committed criminal damage."

Offence level statements assert conclusions about criminal responsibility and liability, which are paradigmatically questions for the court. Expert witnesses should not testify to propositions at the offence level, because they involve factual and moral judgments that forensic scientists are not jurisdictionally competent to make (e.g. Did the victim consent? Was harm caused unlawfully?). Of course, it does not necessarily follow that, in practice, forensic scientists and other expert witnesses are always successful in steering clear of offence level propositions, sometimes there is a trespass beyond the logical scope of their evidence.

3.9 *Practitioner Guide No 4* will present a more systematic analysis of interpretational issues relating to the different levels at which evidential propositions may be stated. For these introductory purposes, it will suffice to underline three fundamental points.

First, it is essential on every occasion to identify the precise question which scientific evidence is being adduced to answer. Testimony offered to answer the question, "What is the source (or even sub-source) of this evidence?" is plainly not equivalent to testimony answering the question, "Did the accused have intercourse with the complainant?", still less does it answer the ultimate question, "Did the accused rape the victim?" Note that these are *all* questions for the fact-finder in criminal proceedings, since they all require inferential conclusions to answer them, albeit at different levels of proposition. That testimony or other evidence is being adduced to answer a particular question does *not* entail that the expert witness should try to answer that question *directly*. Generally speaking, expert witnesses should avoid stating inferential conclusions and instead restrict themselves to commenting on the ***likelihood of the evidence*** under each of two competing propositions, i.e. to expressing and explaining the ***likelihood ratio***.

Secondly, there is a delicate balance to be struck between the transparency and scientific rigour of an expert's evidence and its potential helpfulness to the court. Sub-source propositions are the most rigorous and transparent, but they may not go very far in resolving disputed questions of fact and could be open to misinterpretation (e.g. without guidance, the fact-finder could easily mistake a sub-source level proposition for a source level proposition). Source level propositions, likewise, may have limited utility for criminal adjudication. Even if source level testimony substantially warrants a particular inference, e.g. that a suspect is the source of a blood stain, this does not help determine whether the stains were transferred during a criminal assault or entirely innocently or by a third party. Activity-level propositions come closest to the questions that the fact-finder has to answer, but often build in more speculation and assumptions. The scientist may be able to draw on further relevant expertise, e.g. about transfer and persistence for trace evidence,[17] that can be factored into an activity level proposition and provide valuable assistance to the fact-finder. In every case, however, it is essential that everybody in the courtroom understands the significance of what is being said, that the scientist's assumptions and inferential reasoning should be available to critical scrutiny, and that expert witnesses are able to explain and justify the reasonableness of their assumptions if called upon to do so.

Thirdly, it is worth repeating that evidence evaluation is always a fundamentally comparative enterprise. At all levels of proposition the scientist needs to consider the ***likelihood ratio*** for the evidence, i.e. the probability of the evidence given the prosecution proposition, compared with the probability of the evidence given the defence proposition. Ascertaining the prosecution proposition is normally fairly straightforward, e.g. "the accused is the source of the crime stain at the scene" (paving the way to potential further inferences, that the accused was present at the scene, and that he committed the offence there). It may be more difficult to generate realistic defence propositions if there has been limited pre-trial defence disclosure, although it is always possible to use the negation of the prosecution's proposition as a default setting ("the accused did *not* leave the crime stain at the scene", etc.). Postulating appropriate propositions for comparison is closely tied to the facts of each case, and it is a largely intuitive, non-mathematical exercise, rooted in "logic and experience" (in the sense familiar to criminal lawyers). These important issues affecting the value and interpretation of probabilistic evidence will be further explored and elucidated in *Practitioner Guide No 4*.

---

[17] *R* v *Weller* [2010] EWCA Crim 1085; *R* v *Reed and Reed*; *R* v *Garmson* [2010] 1 Cr App R 23; [2009] EWCA Crim 2698.

3.10 **(b) *Interpreting the language of inferential conclusions***

It is also important to pay close attention to the precise language used in expert reports and testimony to express evidentially significant connections between phenomena (and expert witnesses should correspondingly take care to express such connections precisely). Many forensic scientists and other experts employ stock terminology in report-writing which, although a valid way of expressing preliminary conclusions, may be of limited value to a court and could potentially be misleading unless appropriately qualified and interpreted with circumspection. Further discussion of these ideas may be found in Jackson (2009).

3.11 **"Consistent with"**: It is sometimes said that the evidence is "consistent with" a particular proposition relating to a contested issue in the case, e.g.:

> "Traces of chemicals detected on the swab from the right hand of the suspect *are consistent with* coming from the explosive used at the scene of the explosion."

To say that something is "consistent with" something else means only that the stated proposition (hypothesis) is not excluded by the evidence. It says nothing about how likely the proposition is to be true. For example, buying a ticket is consistent with winning the National Lottery, but it does not make winning very likely. Buying a ticket is also consistent with *not* winning the National Lottery, and this second outcome is very much more likely than the first, though both are equally "consistent with" the premiss (buying a ticket).

3.12 **"Could have come from" / "Could have originated from"**: Once a "match" (however defined) has been obtained between a control sample and a recovered sample, it is common practice for scientists to express an inferential conclusion, such as the following:

- "The semen stain could have come from Mister X, the suspect."
- "The footwear mark at the crime scene could have been made by the shoe the accused was wearing."
- "The blood stain on the window-frame could have been left by the defendant."

- "The fibres recovered from the defendant's clothing could have originated from the victim's sweater."
- "The person shown holding the knife in the CCTV footage could be the defendant."

Statements such as these might be understood as establishing a proven association between the crime and the accused. Notice, however, that "could have come from" does not rule out other possible sources. Indeed, it does not even say that the identified source is the most likely candidate. There may well be other explanations that have not been offered, or even considered, by the scientist, including explanations with a higher probability than the association specified in each statement. Like expressions of "consistency", variations on "could have come from" or "could have originated from" give absolutely no indication of the likelihood that the postulated source is the actual source of the evidence.

3.13 **"Cannot be excluded"**: Another phrase commonly employed by expert witnesses is "cannot be excluded", as in the following examples:

- "The defendant cannot be excluded as the stain donor."
- "The victim cannot be excluded as the source of the blood spatter on the accused's shirt."
- "The broken window cannot be excluded as the source of the glass in the defendant's shoe."

"Cannot be excluded" is the mirror-image of "could have come from" in its vagueness, and is equally susceptible to misinterpretation. There may be any number of alternative sources or explanations that likewise "cannot be excluded", and some of these might be much more likely. The fact that a postulated source cannot be excluded does not mean that evidence of association is strongly or even more than minimally probative.

3.14 A particular variant of the "cannot be excluded" formula is common in DNA and paternity cases, where it is expressed as ***the probability of exclusion***. This probability states what proportion of the population the characteristic would exclude, regardless of who is the donor of the crime-stain. For example, if a relevant characteristic is shared by 0.1% (a relative frequency of 0.001 or 1/1,000) of the population, then the probability of exclusion is 0.999. If a characteristic is present only in 0.1% of the population then it is absent in 99.9% of the population. Thus, if the characteristic is present at the scene of the crime and identified as coming from the (unidentified) perpetrator, 99.9% of the population are excluded as donors of the characteristic.

The probability of exclusion answers the question: "How likely is this characteristic to exclude Mister X if he is *not* the donor of the stain?" However, this could be a very misleading way of expressing the probative value of the evidence, because the court is normally interested in a completely different question: "How much more likely is the evidence if Mister X *is* the donor of the stain than if some randomly selected person were the donor?" (i.e. ***the likelihood ratio***). The probability of exclusion does not address this second, forensically salient question, the answer to which turns crucially on the size of the suspect population. If the relevant population is, say, 1 million, there will be 1,000 individuals with the relevant characteristic, notwithstanding a probability of exclusion of 99.9%.

3.15 Misinterpretations of the probability of exclusion set the pattern for most of the other recalcitrant reasoning errors identified in this section. The trump card, in every case, is scrupulous attention to the meaning of a particular proposition Always ask: what question does this evidence purport to answer? On what assumptions is this statement of probability conditioned? Avoiding elementary probabilistic reasoning errors is as banal and intensely difficult in practice as that.

3.16 **(c)** ***Illegitimately transposing the conditional ("the prosecutor's fallacy")***

Several references have already been made to the probabilistic reasoning error popularly known as "the prosecutor's fallacy", but more technically and accurately described as illegitimately ***transposing the conditional***. This is an error that in principle any participant in criminal proceedings could make: lawyers, judges, jurors, or forensic scientists. In many ways, forensic scientists who fall into this error could be regarded as the chief culprits, since if the expert makes a transpositional error in their initial report or testimony it is eminently foreseeable that lawyers, judges and fact-finders will simply adopt and perpetuate it. After all, the expert is supposed to be the expert and lawyers, judges and lay fact-finders claim no special expertise in reasoning with probabilities. However, erroneous transpositions of the conditional have repeatedly been exposed in scientific evidence – especially DNA profiling testimony – adduced by the prosecution, and illegitimately transposing the conditional has for this reason widely come to be known as "the prosecutor's fallacy". Although not truly apt, the label has stuck.

We saw in Section 2(f) 2.25-2.31, above, that Bayes' Theorem transposes the conditional *legitimately* by employing a valid mathematical formula for this purpose. We are now concerned with evidential propositions which purport to transpose the conditional illegitimately, without employing Bayes' Theorem or any other recognised method of producing a valid conclusion. The error is typically perpetrated unconsciously, and is consequently all the more insidious and liable to precipitate miscarriages of justice for being hidden even from those ostensibly best equipped to avoid it.

3.17 The most direct way of conceptualising the error is to say that it confuses ("transposes") the conditioning event. Consider the following two propositions:

> #1: If I am a monkey, I have two arms and two legs.

> #2: If I have two arms and two legs, I am a monkey.

These conditional propositions ("if….") are clearly not equivalent![18] Proposition #1 is true, whereas proposition #2 is false. Moreover, proposition #2 patently does not follow from proposition #1. When criminal justice professionals illegitimately transpose the conditional they perpetrate an error equivalent to treating proposition #1 as though it were the equivalent of, or at least an authorised version of, proposition #2.

3.18 Utilising shorthand probabilistic notation, the last example can be expressed as follows:

$$p(A+L|M) \neq p(M|A+L);$$

---

[18] Another example of patently non-transitive conditional propositions: #1 "If I am reading this Guide, I can read English"; #2 "If I can read English, I am reading this Guide".

i.e. the probability of Arms and Legs, given (assuming; conditioned on) Monkey is not equal to the probability of Monkey, given (assuming; conditioned on) Arms and Legs.

In the context of criminal proceedings, the standard form of the error 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, i.e.

$$p(E|I) \neq p(I|E).$$

Mathematical notation is particularly useful here, because we can see that "E" and "I" have changed places. On the left hand side of the equation, the conditioning event is "I" ("assuming innocence"). On the right hand side of the equation, "I" has swapped places with "E", which has moved to the left side of the bar indicating the conditioning event ("assuming the evidence"). The conditional has been transposed. These are absolutely not equivalent expressions, as indicated by the "does not equal" sign ($\neq$) dividing the equation.

We have repeatedly stated that the value of evidence is always conditioned on particular assumptions, which should be specified. Consider the following pair of questions about the value of evidence:

> *Assuming that the accused is innocent*, what would be the probability of finding this trace evidence on him?
>
> *Assuming that this trace evidence has been found on the accused*, what is the probability that he is innocent?

The italicised part of each question is the assumption on which the relevant probability is conditioned. The conditional is illegitimately transposed in criminal adjudication when questions of the first type are misrepresented or misinterpreted as questions of the second type.

3.19 A more elaborate illustration should help to make these abstract propositions more readily comprehensible.

Suppose that the DNA profile of a suspect matches the DNA profile from a blood stain found at a crime scene. Assume that the DNA profile has a relative frequency in the *relevant population* of 1/1,000, i.e. one in every thousand people in that country has a matching DNA profile. Let us also stipulate that the relevant *suspect population* (specified through other, non-probabilistic, considerations such as geographical proximity and opportunity) contains 10,001 individuals, the offender and 10,000 innocent others.

One member of the suspect population has been arrested, swabbed, and found to have a DNA profile that matches the profile of the crime stain. Since the relative frequency of the DNA profile in the general population is 1/1,000, the *expected* number of suspects with matching profiles is 10,000 x 1/1,000 = 10. These would be entirely random or "adventitious" matches with entirely innocent individuals. If the offender is known to be the 10,001st member of the suspect population, there are an expected 11 people in the suspect population with matching profiles – ten (expected) "random matches" plus the one offender. It should be emphasised that this "expected" number is a probabilistic projection, not an empirically-observed frequency. Eleven matches are "expected" in exactly the same sense as the "expected" number of heads in ten tosses of a fair coin is five.

3.20 Having been told that the relative frequency of the DNA profile in the general population is 1 in 1,000, it is tempting to equate this to the probability that the suspect is innocent. In other words, to consider the probability of guilt to be 999/1,000; or in notational shorthand, p(Innocent) = 1/1,000; p(Guilty) = 1 – 1/1000 = 999/1,000. But this involves illegitimately transposing the conditional! The stated frequency of 1/1,000 does not represent the probability of the suspect's innocence, but rather the probability that a person picked at random from the general population would have a matching profile, *irrespective of any connection to the offence*.

There are 10,001 people in our suspect population. A particular suspect has been found to have a profile which matches the profile of the crime stain. If the matching profile were the only evidence available, the probability of the suspect's being innocent would be 10/11, which implies p(Guilty) = 1 – 10/11 = 1/11, or 0.09. A probability of 0.09 is not even close to proof on the balance of probabilities, let alone proof beyond reasonable doubt. Yet the error of transposing the conditional produced a fake p(Guilty) of 999/1000 = 0.999, which would easily constitute proof beyond reasonable doubt according to most commentators and participants in empirical research (always allowing for the fact that the courts resolutely refuse to quantify the criminal standard of proof, doubtless for good reason). This stylised illustration demonstrates just how devastatingly powerful such a reasoning error could be in lending credibility to unwarranted conclusions and possibly contributing towards miscarriages of justice.

3.21 Some real-world examples of criminal appeals in which the conditional was illegitimately transposed at trial are given in Appendix B. The so-called "prosecutor's fallacy" tends to be associated with DNA evidence. This is understandable inasmuch as DNA evidence involves quantified probabilities which are articulated in court as random match probabilities, thus routinely presenting opportunities for communication breakdown of one kind or another potentially involving illegitimate transpositions of the conditional.

However, it cannot be stressed too strongly or too often that illegitimate transpositions are *not* a peculiar feature of DNA evidence, but rather potentially could infect every type of evidence, including in particular all kinds of scientific and other expert evidence adduced in criminal proceedings. This follows from the fact that all types of evidence can be assigned subjective probabilities (taking account of relevant data, where available). For example, an expert might testify that there is an 80% probability that mud recovered from the accused's car came from the riverbank near where the deceased's body was discovered;[19] or that there is a "distinct possibility" (perhaps 40%) that handwriting on a forged cheque is the accused's.[20] It would obviously be a crass error to misinterpret these probabilities, respectively, as "an 80% chance of guilt" or "a 40% chance of guilt" of the offences charged. However, both these illustrations of expert testimony involve a more insidious variant of illegitimate transposition, which is described in the next section. The general lesson is that the conditional may be illegitimately transposed *whether or not the evidence is explicitly quantified* and *whether or not expert witnesses realise that they are implicitly drawing upon or assuming probabilistic calculations*.

3.22 **(d) *Source probability error***

[19] *R* v *Shillibier* [2006] EWCA Crim 793, [71].

[20] *R* v *Bilal* [2005] EWCA Crim 1555, [7] – [8].

When illegitimate transpositions of the conditional occur in relation to source level propositions, this is more technically known as *source probability error*.

3.23 Suppose that a crime has been committed, and trace evidence is recovered linking a suspect to the scene, e.g. a DNA match between blood from a murder victim and blood recovered from the suspect's clothing. A scientist determines a value for the frequency of the DNA profile in a relevant population as 1 in 7 million, and writes a report stating:

> "The probability that the blood on the clothing of the suspect came from someone other than the victim is 1 in 7 million. This implies that, with a complementary probability of 6,999,999/ 7 million, the blood on the suspect's clothing came from the victim."

The stated conclusions are unwarranted. They comment erroneously on the source of the blood recovered from the suspect's clothing. It would be legitimate for the scientist to say that, *if* the blood on the clothing of the suspect did not come from the victim, there would be a 1 in 7 million probability of matching the victim's DNA profile. But this is not a proposition about the likelihood of the source; it is the ***random match probability***. In order to calculate the probability that the victim is the source of the blood it would be necessary to know the size of the relevant population (and possibly much else besides, e.g., the probability of an error in testing or of contamination of samples). If there were, say, 14 million potential blood-donors in the relevant population (and making the simplifying assumptions that there is no other pertinent evidence in the case and that all 14 million potential donors were antecedently equally likely to be the true source), the probability that the matching blood came from the victim would be 1/3 (the real victim plus the two other "expected" random matches in the population).

The scientist in this example has transposed the conditional between p(finding a match, assuming the blood on the suspect's clothing could have come from anybody in the relevant population) and p(the blood on the suspect's clothing came from a source other than the victim, assuming a match), i.e. p(Match | Innocent Source) ≠ p(Innocent Source | Match). The scientist then correctly calculates that p(Victim's DNA | Match) = 1 – p(Innocent Source | Match), but irreparable damage has already been done by the initial illegitimate transposition of the conditional. On our assumed frequencies of occurrence in the relevant population, 1 – p(Innocent Source | Match) = 1 – 2/3 = 1/3; again, nowhere near the erroneously asserted value for p(Victim's DNA | Match) of 6,999,999/7 million.

3.24 Returning to the non-DNA illustrations mentioned at the end of the last section (and ignoring for these purposes any complications regarding what constitutes "a match"), the probability that the mud has a common source in the first example is *not* 80%; and the probability that the handwriting in the second example is the accused's is *not* 40%. Rather, these probabilities represent the probability that the recovered sample matches the control sample, *assuming a common source*: p (M | S). To calculate the probability of a common source, p (S | M), it is necessary to factor in the probability that the samples would match, even match perfectly, notwithstanding different sources. Simply put, there could be several – or many – people in the world with identical handwriting, and there could be several – or many – riverbanks with identical mud, just as there may be more than one person in the world with the same DNA profile.

One only needs to mention these possibilities to indicate the difficulties that may be encountered in identifying suitable databases from which to generate reliable frequencies of occurrence for identical handwriting, chemically indistinguishable mud, etc. Setting those complications to one side, we can see that the version of illegitimately transposing the conditional known as source probability error can be, and perhaps frequently is, perpetrated in relation to a range of quantified and unquantified scientific and other expert evidence adduced in criminal proceedings.

The essential insight can be stated as a matter of logic without invoking any formal aspects of mathematics or probability calculations. A measure of similarity or "matching" simply cannot be equated with the likelihood of a common source.

3.25 **(e)** ***Underestimating the value of probabilistic evidence***

Illegitimately transposing the conditional typically makes the evidence in question appear stronger than it actually is. When it relates to prosecution evidence (as it frequently does), illegitimately transposing the conditional constitutes phoney proof of guilt, eroding and potentially undermining the presumption of innocence. There is, however, a complementary reasoning error which involves *under*valuing probabilistic evidence. This was dubbed "the defence attorney's fallacy" by Thompson and Schumann (1987), as a counterpoint to "the prosecutor's fallacy". Again, this terminology is not entirely apt and could mislead, because any participant in litigation, not only defence lawyers, might, in principle, undervalue evidence in this way. Moreover, "the defence attorney's fallacy" is not a true mathematical fallacy (as the so-called prosecutor's fallacy undoubtedly is), but rather a – conceptually speaking – straightforward misrepresentation of the value of probabilistic evidence.

3.26 Suppose that the frequency of blood type AB in a relevant population of 200,000 people is 1%. A suspect is found to have this blood type, matching blood recovered from a broken window at the scene of the crime. Intuitively, this is cogent – albeit not compelling – evidence linking the suspect to the crime-scene.

However, a sceptic might want to argue that the evidence has minimal probative value. The argument supposedly supporting this conclusion runs as follows. There are 200,000 potential suspects, and 2,000 of them would be expected (in the probabilistic sense) to have the blood type AB. If the suspect is merely one of 2,000 similarly situated individuals, the blood evidence might not be thought particularly probative against this, or any other, individual suspect. Indeed, it might now be argued that the evidence is insufficiently probative even to cross the minimal threshold of relevance to warrant legal admissibility. The evidence, it might be said, "proves nothing".

3.27 Although "relevance", "probative value", and "proof beyond reasonable doubt" are indubitably different concepts that need to be carefully distinguished, both in theory and in practice, the sceptical conclusion is overstated. The figure 1/2,000 does *not* represent the value of the evidence of the matching blood type. It is perfectly true to say that, taken in isolation, the blood evidence (merely) places the suspect in a pool of 2,000 potential suspects. However, prior to obtaining the blood evidence the accused was in an undifferentiated pool of 200,000 suspects. The effect of the blood typing evidence is to narrow down that pool by a factor of 100, or in other words to increase the probability in favour of guilt by a factor of 100. Properly evaluated, the evidence is slightly over 100 times more likely if the suspect *is* the source of the blood on the broken window than if he is *not* the source (the probability of a match if the suspect is not the source is 1,999/200,000, or approximately 1/100). In summary, the figure of 100 is taken to represent the value of the evidence. This is powerful evidence, as we intuitively grasp.

Although it would not be capable of proving guilt beyond reasonable doubt if considered in isolation, its probative value is not fairly expressed by saying that the evidence "proves nothing". This interpretational error would be compounded if it were argued, more extravagantly still, that evidence of this kind should be excluded because it lacks sufficient probative value even to qualify as relevant evidence.

3.28 Proof of guilt is normally established, when it is, through a combination of different pieces of incriminating evidence. In Scotland, this expectation is formalised by a formal corroboration requirement necessitating independent evidence of the accused's guilt. Hence, the ultimate value of any particular piece of evidence, scientific or otherwise, must always be assessed contextually, in the light of its contribution to the case as a whole. This general precept is exemplified by the model jury direction suggested by the Court of Appeal in the well-known case of *Doheny and Adams*:

> "Members of the jury, if you accept the scientific evidence called by the Crown, this indicates that there are probably only four or five white males in the United Kingdom from whom that semen could have come. The defendant is one of them. If that is the position, the decision you have to reach, on all the evidence, is whether you are sure that it was the defendant who left that stain or whether it is possible that it was one of the other small group of men who share the same DNA characteristics".[21]

An unusual forensic application described in Gastwirth (1988), drawing on Usher and Stapleton (1979), arose in the following case.

> S, aged 16, became pregnant whilst a patient at a residential facility for those with severe mental disabilities. The pregnancy was terminated and the foetus examined to verify the most likely period of conception and to make serological tests. Because of the limited number (36) of men who possibly could have had access to S and the fact that about 90% of all men could be excluded based on appropriate tests, all 36 were asked to submit to serological tests and all agreed.

---

[21] *R* v *Doheny and Adams* [1997] 1 Cr App R 369, 375, CA. Also see *R* v *Lashley* (2000) and *R* v *Smith* (2000), discussed by Redmayne (2001: 74).

The results of the test excluded all but four men and a further enzyme test excluded one more, reducing the potential list of suspects to three. These three included the police's prime suspect and another other two men regarded as "highly unlikely" to be the perpetrators. The prime suspect was another patient in the home whose disability was somewhat less severe than S's. The principal evidential value of the blood tests in this case was the elimination of innocent men from the list of suspects.

3.29 **(f)** ***Probability ("another match") error***

Two further quantities that are often confused in probabilistic reasoning are: (1) the frequency of an event within a designated population; and (2) the probability of a random match. This error might also potentially infect probabilistic evidence adduced in criminal proceedings, or its interpretation in criminal adjudication. It can be elucidated through a series of simple illustrations.

3.30 Suppose that a crime is committed, and evidence of a blood stain with a profile frequency of 1 in a million is found at the scene and identified as belonging to the offender. Consider the proposition that the evidence was *not* left by a particular suspect.

We know that the frequency of the profile of the stain is 1 in a million amongst the relevant population to which the offender is believed to belong. This means that if a person were chosen at random from that population the probability of that person's profile matching the profile of the blood stain is 1 in a million. This is the ***random match probability***. Notice, however, that this is *not* the same as saying that "the probability of finding another person in the population who has the same genetic profile is 1 in a million". In the first scenario, a person is chosen at random and a DNA profile obtained. The conclusion states the probability of achieving a match "in one go" (akin to the probability of choosing the ace of spades when making one draw from a shuffled standard deck of cards, i.e. 1/52). The second, "another match" probability relates to the occurrence of the event across an entire population, which for the ace of spaces in a standard deck is 1 (the card is definitely somewhere in the pack).

Consider a population comprising one million + 1 individuals, where the additional "+1" is the offender and there are one million innocent people. Then it can be calculated mathematically (see Appendix B) that the probability of at least one match with the offender amongst the one million innocent people is just over 3 out of 5 (0.63). This probability is obviously much larger than the profile frequency of 1 in a million.

3.31 The probability ("another match") error arises when the profile frequency of 1 in a million is equated to the probability of finding at least one other person in the population with the same frequency. A small value for the (random match) profile frequency is taken to imply a small value for the probability that at least one other person has the same matching profile. There is only one chance in 1 million that a person picked at random from the population shares a DNA profile that is common to one in every million people, but there is a 63% chance that there is at least one other person, somewhere, in a population comprising 1 million people who shares that profile.

This result is somewhat counter-intuitive, but it is plainly demonstrable. Consider a "population" of two fair coins, in which for each coin the probability of a head when the coin is tossed is p(head) = 0.5. The coins are secretly tossed once each; we do not know the outcome. Call a third coin, lying heads up, the *crime coin*. The issue is, will the population of tossed fair coins contain a match for the crime coin? The probability of observing at least one coin with a head ("another match") in the tossed coin "population" is not 0.5 (the random match probability for each coin), but 0.75. There are four, and only four, possible outcomes across the tossed coin "population": (a) the first coin is a head, the second coin is a head; (b) the first coin is a head, the second coin is a tail; (c) the first coin is a tail, the second coin is a head; (d) the first coin is a tail, the second coin is a tail. In three out of these four scenarios (75%, or 0.75) there is at least one head, matching the crime coin, in the tossed coin population. Only in scenario (d) is there no matching "head", giving a complementary probability of p(no match with crime coin) equal to ¼ = 25% = 0.25.

Analogously for the DNA profile example, probability ("another match") error is thinking that the probability of finding another person in the population of 1 million people (or 1 million secretly tossed coins) with the same genetic profile as the offender (crime coin) is 1 in a million. But the random match probability figure of 1 in a million is akin to the expected probability of tossing one coin and getting a head (0.5), as opposed to the probability of finding another person (tossed coin) in the population who has the same genetic profile (came up heads) as the offender (crime coin).

3.32 **(g)** ***Numerical conversion error***

Consider a characteristic which is prevalent in only 1 in a thousand, 1/1,000, people (e.g. a height greater than a certain designated value, such as two metres). It is sometimes claimed that the significance of evidence of this characteristic can be expressed in terms of the number of people who would have to be counted before there is another (random) match, being the reciprocal of the frequency (1,000, in this example); i.e. 1,000 people would need to be observed before someone else of that height would be encountered. This is an obvious fallacy, since the very next person observed could be that height or taller.

A frequency of 1/1,000 does not mean that a match (with heights, as in this example, or with any other designated characteristic) is expected only on every thousandth experimental observation. This would almost be like saying that, if one in every thousand motorists will cause a serious accident, we should confiscate the licences of every thousandth driver we encounter. Numerical conversion error featured in the American case of *Ross* v *State*.[22] The relative frequency of a DNA profile was calculated as 1 in 23 million. On the strength of this calculation, the expert testified that he would not expect to encounter another individual with that profile until testing at least another 23 million people. This considerably exaggerates the probative value of a matching DNA profile. It can be calculated mathematically that, for a relative frequency of 1 in 23 million, just under 16 million people would need to be tested in order to achieve a probability of at least 0.5 ("as likely as not") of identifying someone other than the defendant with that profile.

3.33 **(h)** ***False positive probability (distinguished from the probability that a declared match is false)***

[22] *Ross* v *State*, Court of Appeals of Texas, Houston (14th Dist.) 13 February 1992, transcript quoted by Koehler (1993: 34).

Serious errors of interpretation can occur through ignorance or underestimation of the potential for a ***false positive***. A false positive result in a scientific or medical test, for example, 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. In contrast, a ***false negative*** result is one in which the test gives a negative result indicating the absence of the substance or disease, etc. for which the test was conducted when in fact the substance or disease is present.

Many types of scientific and other expert evidence adduced in criminal proceedings have the potential for generating false positives (and false negatives). For example, a forensic scientist might declare "a match" between a DNA profile taken from a crime scene and a DNA profile from a suspect. Suppose, in reality, the suspect does not have the same profile as the perpetrator nor is he the source of the crime scene stain. The result is a false positive. Reported matches relating to fingerprints, ballistics, and various forms of trace evidence (blood, semen, hairs, fibres, firearms residue, etc.), amongst others, are likewise susceptible to false positives (reported matches, where there is no match in fact). The ***false positive probability*** is the probability of reporting a match when the suspect and the real perpetrator do not share the same DNA profile, or where the suspect's and crime-scene fingerprints, blood, fibres or whatever do not, in fact, match.

3.34 Once again, it is vital to pay close attention to the precise wording of these expressions (that is, to specify the precise question which the evidence is being adduced to answer) and to be on one's guard against illegitimate conflations of quite different quantities. Here, in particular, it would be fallacious to equate a value for the ***false positive probability*** (the prior probability of declaring a match falsely) with the value for the ***probability of a false match*** (the probability that any given *declared* match is false). Despite the linguistic similarity of these formulations, they represent categorically different concepts of probability. The first value is a measure of the reliability of testing procedures, which is given by the percentage of non-matches reported as matches (the frequency of reported matches that are not true matches); the second value is the probability that, a match having been declared, it will be a false match. The probability of a false positive is the probability of a match being reported under a specified condition (no match). It does not depend on the probability of that condition occurring, since the condition (no match) is already assumed to have occurred. By contrast, the probability that the samples do not match when a match has been reported depends on *both* the probability of a match being reported under the specified condition (no match) *and* on the prior probability that that condition will occur. Consequently, the probability that a reported match is a true match or a false match cannot be determined from the false positive probability alone.

The distinction between false positive probability and the probability that a declared match is false has important implications for interpreting the reliability and probative value of scientific evidence. A particular laboratory may have a low false positive rate in the sense that it does not often report false matches. However, this does not necessarily mean that when the laboratory declares a match there is a high probability that it is a true match rather than a false positive. The probability that a declared match is a false positive is partly determined by pertinent base rates, which can have unanticipated effects (as we saw in the Blue and Red Buses hypothetical discussed in §2.30–§2.31). The following pair of hypothetical illustrations should serve to reinforce the message.

3.35 Suppose that, in a relevant population of 10,000 individuals, the base-rate for Disease X is 1% (100 people). A person chosen at random from the population therefore has a probability of 0.01 of being infected. The probability that a particular diagnostic test for the disease will give a positive result if a person has the disease is known to be 0.99. So for the 100 people that actually have the disease, 99 will give a positive test result. A negative result would be recorded for the other infected individual, who is the one ***false negative***.

The probability that this same diagnostic test will give a negative result if a person does not have the disease is stipulated to be 0.95. Thus, for the 9,900 people who do not have the disease, 9,405 would give a negative test result. The other 495 people will test positive, even though they do not actually have the disease. They are ***false positives*** and the ***false positive probability*** is 0.05 (5%). Employing the terminology of "sensitivity" and "specificity" introduced in §2.21, we can say that the sensitivity of the diagnostic test is 0.99, and its specificity is 0.95.

These results are summarised in the following table:

Table 3.1: Results of a Diagnostic Test for Disease X

	Diagnostic Test		Total
	Positive	Negative	
Disease X present	99	1	100
Disease X absent	495	9,405	9,900
Total	594	9,406	10,000

Suppose that an individual tests positive for Disease X. What is the probability that this person actually has the disease?

From the table, we can clearly see that the number of people expected to test positive for the disease is 594. Of those 594 people, 99 will actually have the disease. Thus, the probability that a person with a positive result for the test actually has the disease is 99/594 = 1/6. Complementarily, the probability that a person with a positive test result does not have the disease is 495/594 = 5/6.

The diagnostic test is both highly sensitive and highly specific to Disease X, generating an intuitive expectation that the test should be highly reliable. However, because the base rate for the disease in the population is very low (1%) the probability of a declared match being false is surprisingly high – 495/594 = 5/6. The probability that a declared match is a false positive is completely different to the ***false positive probability*** for the diagnostic test, which is a measure of the test's specificity. From the table, we can see that the test will incorrectly diagnose 495 out of the 9,900 people in the population who are not infected with Disease X, i.e. 495/9,000 = 0.05; which is the complement of the test's stipulated specificity (0.95). The probability that a declared match is false varies with changes in the base rate (and at the limit, if the base rate were zero the probability that a declared match is false would be 1, and vice versa), whereas the specificity of a diagnostic test is unaffected by changes in the base rates for infection.

3.36 A second hypothetical example using the same numbers but this time referring to DNA evidence will clarify the significance of this distinction for criminal proceedings.

Table 3.2: Results of DNA Profiling

	DNA Evidence		Total
	Present	Absent	
Guilty	99	1	100
Innocent	495	9,405	9,900
Total	594	9,406	10,000

Consider Table 3.2. In this variation, the prior probability of guilt (base rate) is 1% (100/10,000); the probability that the evidence is detected on a person who is guilty is 0.99 (99/100); the probability the evidence is absent on a person who is innocent is 0.95 (9,405/9,900). The number of people on whom the evidence is present is 594, of whom 99 are guilty. The other 495 on whom the evidence is detected are innocent false positives. Thus, the probability that person on whom the evidence is detected is guilty is 99/594 = 1/6.

The *false positive fallacy* (Thomson *et al* 2003) is to equate the antecedent probability of a false positive (presence of the evidence when a person is innocent) with the probability that a person on whom the evidence is present is nonetheless innocent. In this illustration:

(i) the probability of a false positive is 495/9,900 = 1/20 = 0.05 (in other words, the test is 95% specific for matching DNA profiles);

(ii) the probability a person is innocent when the evidence is present (a match has been declared for the DNA profiles) = 495/594 = 5/6 = 0.833 (approx.).

The second probability is obviously much larger (and the corresponding event more likely) than the first, and it would be a serious error to confuse them with each other.

3.37 **(i) *Fallacious inferences of certainty***

A very low probability of a random match is sometimes thought to equate to a unique identification. For example, a DNA profile with a very small random match probability might be taken to imply that the possibility of encountering another person living on earth with the same DNA profile is effectively zero; in other words, that there is sufficient uniqueness within the observed characteristics to eliminate all other possible donors in the world. Influenced by similar thinking, the US Federal Bureau of Investigation decided that FBI experts could testify that DNA from blood, semen, or other biological crime-stain samples originated from a specific person whenever the random match probability was smaller than 1 in 260 billion (Holden, 1997).

3.38 However, all such inferences of uniqueness are epistemologically unwarranted. Probabilistic modelling must be adjusted to accommodate the empirical realities of criminal proceedings. For example, there may be contrary evidence, such as an alibi, or risks of contamination of samples, etc. Also, some of the modelling assumptions underpinning the probabilistic calculations may be open to challenge. In the final analysis, no probability of any empirical event (e.g. the probability of another person matching a DNA profile), however small, can be equated to a probability of zero (no person with a matching profile living anywhere in the world). Even though a random match probability may be extremely small (one in ten billion, say – the world's estimated current population being (only) six billion) it does not warrant the inference that a matching DNA profile uniquely identifies an individual. Quite apart from anything else, every set of identical twins in the world has the same DNA profile – and the chances of obtaining random matches are vastly increased in relation to parents and siblings.

With a random match probability of, e.g., one in ten billion and a world population of six billion, the probability that there is at least one other person with the profile is about 0.46 (and a corresponding probability of 0.54 that no-one else does). For six billion people and a random match probability of 1 in 260 billion, the probability of at least one other match in the population is about 0.02.

3.39 There appears to be growing sophistication in probabilistic reasoning across the forensic sciences, which has been spearheaded by developments in DNA profiling. Commenting on this trend, Saks and Koehler (2005) anticipate "a paradigm shift in the traditional forensic identification sciences" suggesting that "the time is ripe for the traditional forensic sciences to replace antiquated assumptions of uniqueness and perfection with a more defensible empirical and probabilistic foundation". The idea here is that DNA evidence and the probabilistic techniques applied to it will become a kind of "gold standard" for all forensic science evidence. DNA evidence will be explored at greater length in *Practitioner Guide No 2*.

3.40 **(j) *Unwarranted assumptions of independence***

Probabilistic concepts of independence and dependence were introduced in Section 2 of this *Guide*. Our final "trap for the unwary" involves assuming that two probabilities are independent, and therefore amenable to the ***product rule for independent events***, when that assumption is unwarranted. Either known information demonstrates that the two events are related, or there are insufficient data to make any reliable assumption either way (and therefore the default assumption should be dependence in criminal proceedings).

3.41 A real-world illustration of fallacious assumptions of independence is afforded by *Sally Clark*'s case.[23] Research data showed that the frequency (probability) of sudden infant death syndrome (SIDS) in a family like the Clarks' was approximately 1 in 8,543. From this it was deduced, applying the product rule for independent events, that the probability of two SIDS deaths in the same family would be 1/8,543 x 1/8,543 = 1/72,982,849, which was rounded down to produce the now notorious statistic of "1 in 73 million" quoted in court. The fact-finder was apparently encouraged to believe that the figure of 1 in 73 million implied that multiple SIDS deaths in the same family would be expected to occur about once every hundred years in England and Wales. Of course, this calculation and deduction are valid only on the assumption that two SIDS deaths in the same family are entirely unrelated, independent, events. But this was a perilously fallacious assumption.

In reality, the assumption of independence was directly contradicted by the research study from which the original 1/8,543 statistic was derived. Fleming *et al* (2000) reported that a sibling had previously died and the death ascribed to SIDS in more researched SIDS families than in control sample families (1.5%, five out of 323 families, and 0.15%, two out of 1288 families, respectively, and that these percentages were significantly different in the statistical sense). Far from warranting an assumption of independence, these empirical data suggest that multiple SIDS in the same family may be dependent events.

3.42 Recall that interpretation of evidence is a fundamentally comparative exercise. The true probative value of evidence can be assessed only by considering it under at least two propositions, which in criminal proceedings can be modelled as "the proposition advanced

[23] *R* v *Clark* [2003] EWCA Crim 1020.

by the prosecution" and the competing "proposition advanced by the defence" (which, in the absence of anything more suitable, may simply be the negation of the prosecution's proposition).

When the evidence is implausible under the defence proposition, it is tempting to jump to the conclusion that the prosecution's case (proposition) must be true. But that inference is speciously premature. The evidence might be *even more* implausible assuming the truth of the prosecution's proposition. For example, it might be very unlikely that two cases of SIDS would be experienced in a single family. But it might be even less likely that a mother would serially murder her two children (we must make assumptions here, of course, about the impact of other evidence). So, taken in isolation, the bare fact of two infant deaths in the same family is probably more likely to be SIDS than murder. Unlikely though the former innocent explanation may be, it is not as unlikely as the latter, incriminating explanation.

3.43 Forensic scientists and other expert witnesses in criminal proceedings should guard against making unwarranted assumptions of independence. That two events or characteristics are truly independent should be demonstrated rather than merely assumed before applying the product rule for independent events to calculate the probability of their conjunction. Witnesses who testify on the basis of independence should be prepared to explain and justify their rationale for that supposition, whilst lawyers should be ready to probe statements of the form "research shows that…" in order to satisfy themselves that the quoted research is fit for purpose and that the evidence does not rest on unwarranted assumptions of independence.

4. Summary and Checklist

4.1 **Introduction:** ***Communicating and Interpreting Statistical Evidence in the Administration of Criminal Justice***

Statistical evidence and probabilistic reasoning place intellectual demands on most of the professional participants in criminal proceedings, including lawyers, judges and expert witnesses. There is no room for complacency; errors and misunderstandings relating to probability and statistics have contributed towards serious miscarriages of justice.

4.2 Every professional participant in criminal proceedings should ideally acquire sufficient knowledge of probability and cultivate the practical competence needed to interpret statistical information correctly in order to fulfil their respective roles in the administration of criminal justice. Probability is one specialised dimension of logical reasoning. Criminal justice professionals may or may not find it illuminating or convenient to employ the formal tools of probability and statistics in their own professional practice, but they do need to be able to recognise these techniques and successfully decode them when they are invoked or implicitly relied on by others. Moreover, the prospect of implicit or unconscious reliance on probabilistic reasoning places an even greater premium on vigilance. In short, judges, lawyers and expert witnesses should be ***responsible producers*** and ***discerning consumers*** of statistical information and probabilistic reasoning whenever they are introduced into criminal proceedings.

4.3 **1. Probability and Statistics in Forensic Contexts**

Statistics are generalisations derived from observations of the empirical world. Statistical reasoning is characteristically inductive. Probability, by contrast, is a way of measuring uncertainty which is projected onto the world and thereby helps us to formulate and implements rational plans of action. Probabilistic reasoning is deductive. Both topics may be regarded as overlapping but conceptually distinct parts of the larger human endeavour of reasoning under uncertainty, of which criminal adjudication is one important manifestation. Probability obeys mathematical axioms with powerful real-world applications, which include important aspects of evidence and proof in criminal proceedings.

4.4 Statistics has many forensic applications, but it must be approached with care and interpreted correctly. There are many equally valid ways of presenting statistical data. For example, the mean, the median, the mode and the standard deviation are alternative ways of summarising estimates which emphasise different aspects of relevant data. The question is not whether these alternative estimates are "right" or "wrong", but rather whether they are suitable for particular purposes. Thus, confidence intervals are regarded as appropriate expressions of uncertainty in social science and elsewhere, but they are not an appropriate way of evaluating evidence in criminal proceedings because they are irremediably arbitrary and unjustifiably cause valuable evidence to "fall off a cliff".

The validity of statistics is a function of sampling techniques and other methodological considerations, which need to be taken into account when assessing inferential conclusion based on statistical information. Probability theory can help with these assessments. In the final analysis, statistical inferences can only be as good (or as poor) as their underlying data.

4.5 In summary, when statistics are being presented and interpreted in forensic (or any other) contexts, there are always two principal dimensions of analysis to be borne in mind:

(1) *Research methodology and data collection*: Do statistical data faithfully represent and reliably summarise the underlying phenomena of interest? Do they accurately describe relevant features of the empirical world?

(2) *The epistemic logic of statistical inference*: Do statistical data robustly support the inference(s) which they are assumed or asserted to warrant? Is it appropriate to rely on particular inferential conclusions derived from statistical data?

4.6 **2. Basic Concepts of Probabilistic Inference and Evidence**

The starting point for the interpretation and evaluation of evidence is to identify the precise question that it purports to answer. More specifically, one must consider:

- How is the evidence relevant? (Irrelevant evidence is never admissible.)

- If relevant, does the evidence fall foul of any general exclusionary rule?
- If admissible, what is the probative value of the evidence?

Insofar as probabilistic evidence and reasoning involve specialist skills and knowledge, legal professionals and expert witnesses should be able to discharge their allotted roles responsibly and in accordance with the interests of justice by mastering a relatively small number of basic concepts, theorems and other applications (such as the product rule for calculating the conjunctive probability of independent events). Probability theory is often illustrated through contrived examples involving tossing coins, drawing playing cards from a normal deck, spinning a roulette wheel, and the like. However, these hypothetical contrivances have powerful real-world implications, not least for criminal adjudication.

4.7 ***Relative frequencies*** provide basic units of probability with the most immediate and extensive forensic applications. As base rates, frequencies relate to general variables or to background data such as production or sales figures. When incorporated into expert reports or testimony adduced in criminal proceedings, frequencies more commonly relate to case-specific evidence. All such relative frequencies informing probabilities are predicated or "conditioned" on certain assumptions. These assumptions should be specified in every case, and their adequacy for the task in hand explored, interrogated and verified.

4.8 Evidence evaluation is always a fundamentally comparative exercise. Ideally, expert witnesses should testify to the ***likelihood of the evidence*** under two competing propositions (or assumptions), the prosecution's proposition and the competing proposition advanced by the defence (which may simply be the negation of the prosecution's proposition in the absence of fuller pre-trial defence disclosure). In other words, experts should testify to ***the likelihood ratio***. 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 two competing propositions.

4.9 ***Bayes' Theorem*** states that the posterior odds are equal to the prior odds multiplied by the likelihood ratio. This theorem authorises legitimate transpositions of the conditional, converting the probability of the evidence assuming guilt – p(E|G) – into the probability of guilt assuming the evidence; p(G|E). Bayesian reasoning applies most directly to quantified evidence, such as DNA profiles with mathematically calculable ***random match probabilities***. However, Bayes' Theorem can in principle be extended to any kind of evidence, since one can always, theoretically, attach subjective probabilities to unquantified evidence of any description. The reasonableness of any subjective probability is always open to question, and its underlying assumptions should be identified and thoroughly tested in criminal litigation. Although the Court of Appeal has denounced attempts to encourage *jurors* to attempt Bayesian calculations, especially in relation to non-scientific evidence, many forensic scientists are confirmed or unconscious Bayesians and routinely employ likelihood ratios in the course of generating expert evidence ultimately adduced in court. This is entirely appropriate and justifiable (Bayes' Theorem is, after all, a valid deduction from mathematical axioms), *provided that such evidence is properly interpreted* and its underlying assumptions and limitations are correctly identified and evaluated.

4.10 Probabilistic evidence of all kinds is susceptible to recurrent reasoning errors. Bayes' Theorem, for example, is associated with the so-called "prosecutor's fallacy". This *Guide* sought to identify, deconstruct and neutralise the most frequently encountered and persistent of these probabilistic "traps for the unwary".

4.11 **3. Interpreting Probabilistic Evidence – Anticipating Traps for the Unwary**

Expert evidence employing probabilistic concepts or reasoning may address different levels of proposition. It is essential to ascertain (and for experts themselves to state clearly) whether the evidence addresses source, sub-source or activity-level propositions. Source and – especially – sub-source propositions afford the most focused and narrowly circumscribed ways of expressing an expert's inferential conclusions, but they are not necessarily the most helpful to the court. Activity-level propositions are generally more helpful in resolving disputed questions of fact but tend to build in more inferential steps and are consequently, in this sense, less transparent regarding their underlying data and conditioning assumptions. In every case, it is the forensic scientist's duty to identify the data and spell out the assumptions on which their expressed opinion is based. Experts should always steer clear of crime-level propositions, which are exclusively reserved to fact-finders in criminal adjudication.

4.12 It is also important to pay close attention to the nuanced language of expert reports. Phrases such as "consistent with", "could have come from" and "cannot be excluded" are potentially misleading, inasmuch as they give no indication of the probative value of an asserted association. In fact, such conclusions are virtually meaningless unless pertinent alternatives are also considered.

4.13 The conditional is illegitimately transposed when the probability of the evidence conditioned on innocence, p(E|I), is confused with the probability of innocence conditioned on the evidence, p(I|E). These are completely different concepts which often have radically different values. Mistaking one for the other is popularly known as "the prosecutor's fallacy" owing to its (contingent) association with prosecution evidence, especially DNA profiling evidence. However, any participant in criminal proceedings – including forensic scientists and other expert witnesses – potentially can, and many frequently do, fall into this notorious trap.

A variant of the illegitimate transposition of the conditional is known as the ***source probability error***. This is perpetrated by confusing the probability of a match when the suspect is not the source, p(Match | Suspect not the source), with the probability the suspect is not the source assuming matching trace evidence, p(Suspect is not the source | Match). The first quantity is the ***random match probability***; the second is predicated on a positive test result and depends on the size of the population of interest. As before, these quantities could represent dramatically different probabilities. A very small random match probability, for example, cannot be equated to a very small probability that matching samples in fact came from different sources.

The conditional is legitimately transposed through the application of Bayes' Theorem. Illegitimate transpositions arise through confusion and are always unjustifiable. Whether replicating the classical "prosecutor's fallacy" or some variation on source probability error, illegitimate transpositions adopt the flawed logic of thinking that "If I am a monkey, I have two arms and two legs" implies that "If I have two arms and two legs, I am a monkey".

4.14 A different kind of interpretative error involves *under*valuing probabilistic evidence. Evidence can be highly probative even if, taken in isolation, it falls a long way short of constituting proof beyond reasonable doubt. Probabilistic evidence should not be disparaged, must less spuriously rejected as irrelevant, just because it fails to constitute self-sufficient and irrefutable proof of guilt. If this were the authentic legal test of relevance and admissibility, no evidence would ever be given in criminal trials.

4.15 Further potential traps for the unwary lurk in the ease with which it is possible to confuse different probabilities or inadvertently break the axiomatic laws of probability. The following are particularly noteworthy and demand constant vigilance:

- The ***random match probability*** 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).

- A population frequency does not state the number of items of interest that would need to be tested before a match is found. If there were 1,000 plastic balls in a bag, 999 white and one black, the frequency of black balls in the ball population is 1/1,000 but this clearly does not imply that one would expect to pull a black ball out of the bag only at the 1,000th attempt. Fallaciously equating these quantities is known as ***numerical conversion error***.

- The ***false positive probability*** must not be confused with the probability that a stated match is false. The false positive probability is a measure of the ***specificity*** of the test – with what regularity it produces an erroneous match. The probability that a stated match is false turns crucially on the relevant base rates, which are capable of producing strikingly counter-intuitive results on certain empirically plausible assumptions. Even a test with exceedingly good specificity – e.g. a false positive probability of 0.001 (one in a thousand) – will be wrong on every occasion that it declares a match if there are no true positives in the tested population: i.e. the probability that a declared match is false would be 1.

- No random match probability, no matter how tiny, can warrant any inference with 100% certainty, e.g. a unique identification of a particular individual. Probability is concerned with uncertainty all the way to the vanishing point.

- The ***product rule for independent events*** for calculating conjunctive probabilities should be applied only to verifiably independent events. Independence should never be a default assumption in criminal proceedings, where erroneous inferences risk serious miscarriages of justice. Independence must be demonstrated and verified before the product rule for independent events can safely be applied.

.

Appendix A – Glossary

‘|’, the conditioning bar: the vertical line used, in conjunction with p( ), to express conditional probabilities in mathematical notation. The event to the left of the conditioning bar is the unknown variable of interest for which a probability is to be calculated; the assumed or known event is located to the right of the bar. For example, p(Evidence | Guilt) denotes *the probability of the evidence assuming guilt* (not to be confused with p(Guilt | Evidence), *the probability of guilt assuming the evidence*) .

p( ), probability: Notational shorthand for the probability of the event or other variable in the parentheses. For example, p(G) denotes the probability that the accused is guilty; p(I) denotes the probability that the accused is innocent; and p(E) is shorthand for the probability of the evidence.

***x*:** symbol to denote “event” or other variable of interest. Often used in conjunction with p( ), where p(x) denotes the probability of the variable x.

Absolute frequency, see ***frequency***.

Addition rule of probability: For two mutually exclusive events or characteristics (i.e. their ***conjunction*** is impossible), the probability of one or the other being the case is the sum of the probabilities for each individual event. Thus, for blood groups A and AB, the probability that a person is A or AB is the sum of the probabilities (i) that they are A and (ii) that they are AB, or in notation p(A or AB) = p(A) + p(AB). Where events are *not* mutually exclusive, the probability of one or the other (or both) being the case equals the sum of the probabilities for each individual event or characteristic minus the probability of their conjunction, i.e. p(A or AB) = p(A) + p(B) – p(AB). Thus, the probability of having blue eyes and blond hair equals the probability of having blue eyes plus the probability of having blond hair, minus the probability of having both blue eyes and blond hair.

Base rates, or background rates: The rate of occurrence or proportion of some event in a population of relevance to the matter being investigated. In criminal proceedings, this might be the proportion of shoes of a particular design sold in the local area or during a specified time period, etc; or the number of cars with sliver metallic paint as a proportion of all cars sold in the last five years, or currently on the roads, etc.

Bayes' Theorem: a formula for legitimately "transposing the conditional", according to which the ***posterior odds*** are equal to the product of the ***likelihood ratio*** and the ***prior odds***. For example, the posterior odds in favour of guilt after having heard (*conditioned on*) the evidence is the product of (i) the likelihood ratio of the evidence and (ii) the prior odds in favour of guilt before the evidence was heard. The ***likelihood ratio*** is the ratio of (i) the probability of the evidence *assuming that the prosecution's proposition is correct* to (ii) the probability of the evidence *assuming that the negation of the prosecution's proposition ("the defence proposition") is correct.*.

Census: collection of data from the entire population of interest (in contrast to a "sample" comprising some subset of these data – see ***sampling***).

Complementary events, see ***events***.

Confidence interval: an interval constructed from a sample within which a population characteristic is said to lie with a specified degree of confidence, e.g., a "95% confidence interval". Confidence internals typically describe the sample mean plus or minus a multiple of the standard error (the multiple chosen from the specified level of confidence).

Conjunction: The conjunction of two events, *x* and *y*, is the event defined by the occurrence of both *x* and *y*. Thus, the conjunction of the event 'accused has soil on his shoes' and the event 'shoe tread is similar to footprint in soil outside window of burgled house' is the event 'accused has soil on his shoes whose tread is similar to footprint in soil outside window of burgled house'.

Convexity Rule: For any event or issue, the probability of its occurrence can be expressed as a numerical value between 0 and 1 inclusive. Only impossible events have a probability of zero (***Cromwell's rule***). If a probability of zero is assigned to any issue (such as guilt or innocence) no evidence can ever alter that probability.

Count: the number (n) of times a certain event occurs. This could be the number of children in a family, the number of heads in 20 tosses of a coin, the number of times a ball falls in the '1' slot in a roulette wheel, the number of consecutive matching striations in a bullet found at a crime scene and a bullet fired from a suspect gun, or any variable of interest that can be counted, as distinct from a ***measurement***. Counts are whole numbers (integers), 0, 1, 2, etc. However, the ***mean*** of a set of counts need not be an integer, e.g., the mean number of children in British families could be 1.5. Measurements need not take integer values.

Cromwell's Rule: only impossible events can realistically be assigned a value of zero (referring to Oliver Cromwell's plea to the General Assembly of the Church of Scotland on 3 August 1650: "I beseech you, in the bowels of Christ, think it possible that you may be mistaken"(*Oxford Dictionary of Quotations*, 3rd edn 1979).

Deductive logic, deduction: inferential conclusion, typically involving reasoning from generals to particulars (and contrasted with ***induction***). In the standard deductive syllogism, a deductive conclusion follows by logical necessity from initially demonstrated or accepted axioms or premisses

Dependent events: "events" (or, sometimes, "variables") which affect the probability of some other event (variable) of interest. For example, the probability that an unknown person is male is affected by our knowledge of that person's height, and even more so by knowing their name. Likewise, knowledge of size and shape of tyre marks left at a crime scene affect the probability that the marks were created by a particular make and model of getaway car.

Disjunction: The disjunction of two events, x and y, is the event defined by the occurrence of x or y or x-and-y. The disjunction of the event "the accused has soil on his shoes" and the event "the shoes match the footprint at the crime scene" is the event "the accused has soil on his shoes; *or* the shoes match the footprint at the crime scene; *or* both the accused has soil on his shoes *and* the shoes match the footprint at the crime scene".

Error: as a statistical term, denotes the natural variation in a sample statistic or in the estimate of a population characteristic (see also ***standard error***). Statistical "error" has nothing to do with "mistakes" in common parlance.

Events: states of affairs of interest, about which evidence may be given and probabilities calculated. One might refer to: "the event that the suspect's DNA matches the crime stain sample"; "the event that the chemical composition of drugs from two different seizures is identical"; "the probability of the event that fibres from a crime scene match the accused's jumper", etc. ***Complementary events*** are two events such that one or the other must occur but not both together, i.e. $p(x) + p(y) = 1$. The event that a defendant is factually guilty and the event that a defendant is factually innocent are complementary, since the accused must be one or the other; he cannot be both or neither.

Evidence: information relied on for a particular inferential purpose, such as deciding whether the accused is guilty in criminal proceedings. "Legal evidence", "judicial evidence", and – in its original, literal meaning – "forensic evidence" are all synonyms for information which is admissible as evidence in legal proceedings. The principal forms of legal evidence are witness testimony, written statements, documents and physical objects (the latter are known as "real evidence"). The probative value of the evidence can be expressed in terms of conditional probabilities, i.e. as the ratio of the probability of the evidence conditioned on the prosecution proposition and the probability of the evidence conditioned on the defence proposition.

Experiment: the collection of data in a controlled, (as we say) "scientific" fashion seeking to test a specified hypothesis (e.g. regarding the anticipated impact of particular variables) whilst eliminating potentially confounding factors. In an agricultural experiment, different fertilisers might be applied to different areas of farmland to allow variations in crop yield to be documented and assessed. A forensic scientist might compare the different patterns of glass fragments produced when rocks are thrown at windows from varying distances. Purely observational studies, involving no manipulation or intervention by the investigator, are not experiments in the formal sense, although they are sometimes described as "natural experiments" (and may be the only kind of research possible regarding particular questions, owing to ethical or practical constraints).

Facts in issue, see ***issue***.

False match: a match is declared but the identification is false. This could arise for a variety of reasons, including: (i) faulty criteria for declaring a match; (ii) misapplication of those criteria in practice, e.g. a fingerprint examiner erroneously judges two characteristics to be similar when they are dissimilar; (iii) confusion, contamination, or degradation of samples; or (iv) the crime sample and the control sample genuinely do match, but the accused is not in fact the source of the crime sample.

False negative: a negative test result in a case where the feature being tested for (a disease, a chemical substance, a matching fingerprint, etc.) is actually present.

False positive: a positive test result in a case where the feature being tested for (a disease; a chemical substance; a matching fingerprint, etc.) is not actually present.

Frequency,

absolute frequency (of occurrence): the count of the number of items in a certain class, e.g. the number of sixes in 20 throws of a six-sided die; or the number of times the ball lands in the '1' slot in 1,000 spins of a roulette wheel.

relative frequency (of occurrence): the proportion of the number of items in a certain class, e.g. the proportion of sixes in 20 throws of a six-sided die (i.e. the ***absolute frequency*** of sixes divided by 20); or the proportion of times the ball lands in the '1' slot in 1,000 spins of a roulette wheel (the number of balls in the '1' slot divided by 1,000). Proportions take values between 0 and 1; and the sum of proportions over all possible outcomes (1, 2, …, 6 for throws of a die; 0, 1, 2, .. 36 for a 37-slot roulette wheel) equals 1. Proportions can be converted into percentages by multiplying by 100 (thus, where a six is rolled four times in 20 throws of the die, the relative frequency of sixes is 4/20 = 1/5; which multiplied by 100, equals 20% sixes).

Independence, independent events or variables: events or variables *x* and *y* are "independent" when the occurrence or non-occurrence of *x* has no bearing whatever on the occurrence or non-occurrence of *y*. For example, successive tosses of a fair coin or rolls of a fair die are independent events. Independence is not a general default assumption; one must have good grounds for believing that two variables are genuinely independent. In forensic contexts in particular, it is perilous to apply the ***multiplication rule for independent events*** where assumptions of independence are unwarranted.

Induction: in logic, "[t]he process of inferring a general law or principle from the observation of particular instances" (*OED*, 2nd edn 1989). More generally, induction may involve the formulation of empirically-based generalizations and their application to particular cases.

Issue: the matter under investigation, that which is to be determined. In criminal proceedings, the "***facts in issue***" are defined by the elements of the offence(s) charged and any affirmative defences that the accused might advance. The ultimate issue in a criminal trial is whether the accused had been proved guilty to the requisite criminal standard ("beyond reasonable doubt", or so that the fact-finder is sure of the accused's guilt).

Likelihood ratio: a measure of the value of evidence in terms of two probabilities conditioned on different assumptions. The likelihood ratio is the core component of ***Bayes' Theorem***. In relation to evidence of the accused's guilt, for example, this is the ratio of (i) the probability of the evidence on the assumption that the accused is guilty to (ii) the probability of the evidence on the assumption that the accused is not guilty.

Mean: the average of a set of numbers. The mean is the sum of the numbers divided by the number of members comprising the set.

Measurement: a quantity that can be represented on a continuous line, in contrast to a numerical count which always takes a non-negative integer value (0, 1, 2, etc.). For example, height is a continuous quantity. Other continuous quantities relevant to criminal proceedings include the chemical composition of drugs and the elemental composition of glass.

Measures of dispersion: quantitative expressions of the degree of variation or dispersion of values in a population or sample, e.g. the ***standard deviation***.

Median: the value dividing an ordered data set (one in which the members of the set are given in order of ascending or descending value) into two equal halves. For a set with an odd number of members, the median is the middle value, for a set with an even number of members, the median is half-way between the two middle values.

Mode: the value which occurs most often in a set. If there are two values which occur most often the set is bimodal and if there are more than two such values, the set is multimodal.

Multiplication rule, or product rule: see Appendix B.

for independent events: the probability of x-and-y, where x and y are independent, equals the probability of x multiplied by the probability of y, i.e. $p(x \text{ and } y) = p(x) \times p(y)$.

for non-independent ("dependent") events: the probability of x-and-y, where x and y are *dependent*, equals the probability of x multiplied by the probability of *y given that x has occurred*, i.e. $p(x \text{ and } y) = p(x) \times p(y \mid x)$. This also equals the probability of y multiplied by the probability of *x given that y has occurred*, i.e. $p(x \text{ and } y) = p(y) \times p(x \mid y)$.

Nonprobability convenience samples: see ***sampling, convenience***.

Numerical conversion error: The fallacious equation of the reciprocal of a population frequency with the number of items of interest that would need to be tested before a match is found.

Odds: a way of expressing likelihood or probability, in terms of the ratio of the probabilities of two ***complementary events***, i.e. two events, x and y, that are mutually exclusive and exhaustive (either x or y must be the case, but their conjunction is impossible). The odds in favour of x are then p(x)/p(y). For example, a defendant is factually guilty or factually innocent of the crime with which he is charged, and there is no third option ("neither guilty or innocent"; or "both guilty and innocent"). The ratio of the probability of guilt to the probability of innocence is the odds in favour of guilt (the first named event); or the odds against innocence (the second named event). In sport, we speak of the odds against a horse winning a race or a football team winning a match or a competition. The odds version of ***Bayes Theorem*** incorporates ***prior odds*** and ***posterior odds*** in its formula for ***transposing the conditional***.

Odds ratio: the ratio of two sets of odds. For example, in ***R* v *Clark* [2003] EWCA Crim 1020** a research report calculated the odds in favour of a previous SIDS death amongst the study families selected because of a current SIDS death ("cases") and the odds in favour of a previous SIDS death amongst control families with no current SIDS death. The odds in favour of a previous SIDS death in the case families was 5/318; in the control families the odds were 2/1,286. The ratio of these odds is 5/318 divided by 2/1286, which is approximately 10. This result may be expressed as "the odds in favour of a previous SIDS death amongst case families was about 10 times the odds in favour of a previous SIDS death amongst the control families".

Population,

target: the entire set of individuals or items about which information is sought, in other words the "population of interest".

sampled: the population from which a sample is taken. It is essential to try to ensure that the sampled population is the same as the target population. In a crime involving fibre comparisons, for example, the target population is the population of fibres with which the ***recovered sample*** ought to be compared. Defining an appropriate target population involves contextual judgements which may be open to dispute. The sampled population is the population of fibres against which the recovered sample actually is compared. If woollen fibres are known to come from items of clothing, an appropriate target population might be items of woollen clothing rather than, e.g., carpet fibres.

Posterior probability: employed in ***Bayes' Theorem***, the probability after consideration of specified evidence.

Prior probability: employed in ***Bayes' Theorem***, the probability before consideration of specified evidence.

Probability: is a quantified measure of uncertainty. Some probabilities are objective, in the sense that they conform to logical axioms (e.g. the outcomes of tossing a fair coin or rolling a fair six-sided die). Subjective probabilities, by contrast, measure the strength of a person's beliefs, e.g. in the likely outcome of a sporting event, in the accused's guilt, in a witness's veracity. Subjective and objective probabilities of events can be combined when applying the laws of probability. For example, when applying the ***multiplication rule*** to calculate p(x and y), either p(x) or p(y) could be subjective or objective.

Probability of exclusion: the proportion of a particular population that a specified characteristic would exclude. For example, if one in five people in the UK has blue eyes, the probability that a person chosen at random from this population has blue eyes is 1/5. The probability of exclusion for the characteristic 'blue eyes' is 4/5.

Production figures: data summarising the number of items of a particular kind produced by a specified manufacturer and/or over a specified time period and/or in a specified area. Production figures are sometimes adduced in evidence in criminal proceedings as proxies for ***relative frequency of occurrence***.

Product rule: see ***multiplication rule***

Proposition: in the context of criminal proceedings, an assertion or hypothesis relating to particular ***facts in issue***. The probative value of scientific – or any other – evidence may be expressed in terms of the parties' competing propositions, e.g. "the pattern of blood spatter on the accused's clothing supports the *prosecution's proposition* that the accused repeatedly struck the victim with his fist rather than the *defence proposition* that the accused was merely a bystander who took no part in the assault".

crime level: a proposition about the commission of a criminal offence.

activity level: a proposition about human conduct, which could be "active" such as kicking the victim, breaking a window, or having intercourse; or passive, such as standing still.

source level: a proposition about the source of physical evidence, such as the source of fibres on a shoe, paint fragments on clothes, semen at the crime scene, etc.

sub-source level: a proposition about physical evidence which does not purport to specify its provenance or derivation. This level of proposition may be appropriate where a forensic scientist is unable to attribute analytical findings to specific source material. It is commonly used to express DNA profiling evidence where the profile cannot be attributed to a particular crime stain, tissue sample or other particularised source material.

Prosecutor's fallacy, the: common, if rather imprecise, name for the reasoning error involved in illegitimately ***transposing the conditional***.

Random match probability: the probability that an item selected at random from some population will "match" (in some defined sense of "matching") another pre-selected item. For example, a DNA profile is obtained from a blood stain at the scene of a crime. The random match probability is the probability that the DNA profile of a person chosen at random from the general population will match the profile derived from the crime scene.

Random occurrence ratio: a phrase which some lawyers and courts have used as a synonym for the ***random match probability***. However, this terminology is misleading since the random match probability is not, in fact, a ratio.

Reciprocal: the reciprocal of a number is that other number such that the product of the two numbers equals 1. For example, the reciprocal of 6 is 1/6; the reciprocal of 1/6 is 6; the reciprocal of 25 is 0.04; the reciprocal of 0.04 is 25, etc.

Relative frequency, see ***frequency***.

Sales figures: data summarising the number of items sold by a specified retailer and/or over a specified time period and/or in a specified area. Such data are sometimes adduced in evidence in criminal proceedings as proxies for ***relative frequency of occurrence***.

Samples,

control, **or reference**: a sample whose source is known, such as fragments of glass known to derive from a broken window at a crime scene, fibres taken from an article of clothing under controlled conditions, etc.

crime: a sample associated with a crime scene. This could be a ***recovered sample*** *or* a ***control sample***, depending on the nature of the inquiry being undertaken and the matter sought to be proved.

recovered, or questioned: a sample whose source is unknown, such as fragments of glass found on a suspect's clothing, external (foreign) fibres taken from a crime scene, a footwear mark at the scene of the crime, etc.

suspect: a sample associated with a suspect. This could be a ***recovered sample*** *or* a ***control sample***, depending on the nature of the inquiry being undertaken and the matter sought to be proved.

Sampling,

convenience: a sample which has been taken because random sampling is impossible or impracticable. Also sometimes known as ***nonprobability convenience samples***. Convenience sampling must be carefully controlled and evaluated in order to mitigate the risks of bias in the sample, i.e. the ***sampled population*** may fail to match the ***target population***.

random: a sample in which every member of a ***population*** is equally likely to be selected. This may be facilitated by constructing a list, known as a ***sampling frame***, of all members of the population. Sometimes this task is relatively straightforward, e.g. deriving a sampling frame for an electorate from an electoral register. Other kinds of sampling frame may be difficult or virtually impossible to construct in practice, such as the creation of a list of all beer bottles in order to sample glass from beer bottles.

stratified: ***populations*** may sometimes usefully be divided into sections known as ***strata*** defined by relevant characteristics of interest (e.g. within a population of consumers, those who eat all meats; those who eat only fish and chicken; vegetarians; vegans, etc). A ***stratified sample*** contains suitable proportions from each pertinent stratum of the population. For drug sampling from a collection of plastic bags, the strata could be the plastic bags, and a suitable proportion (sample) of drugs could be taken from each bag (stratum).

Sampling frame: see ***sampling, random***

Sensitivity: a measure of a test's ability to detect the presence of the thing it is supposed to be testing for. In a medical context, this might be the probability of a positive test result if a patient does in fact have the targeted disease. More generally in forensic science, sensitivity is expressed as the probability of a positive test result indicating a common source for ***control*** and ***recovered samples*** if the samples do indeed come from a common source. Sensitivity is to be distinguished from ***specificity*** (a particular test could be highly sensitive but not at all specific, leading to a high proportion of ***false positives***).

Source probability error: fallaciously equating (i) the probability of finding a "match" between a ***control sample*** and a ***recovered sample*** where there is no common source (i.e. the ***random match probability***) with (ii) the probability that two samples do not have a common source, where a "match" has been found.

Specificity: a measure of a test's exclusivity in detecting the presence of the thing it is supposed to be testing for. In a medical context, this might be the probability of a negative test result if a patient does not in fact have the targeted disease. More generally in forensic science, specificity is expressed as the probability of a negative test result indicating that ***control*** and ***recovered samples*** have different sources if the samples do indeed come from different sources. Specificity is to be distinguished from ***sensitivity*** (a particular test could be highly specific but not at all sensitive, leading to a high proportion of ***false negatives***).

Standard deviation: a measure of the variation in a ***sample*** or a ***population***. In a sample, the standard deviation is the square root of the division of the sum of squares of deviations of the observations in the sample from the sample mean by a number one less than the sample size.

Standard error: the standard deviation of a sample, divided by the square root of the sample size. It is a measure of the precision of the sample mean as an estimate of the population mean.

Statistic: a number conveniently summarising quantified data, often presented as a percentage or in graphical form using graphs, bar charts, pie charts, etc. Statistics normally refer to a ***sample*** rather than a ***census***.

Strata, see ***sampling, stratified***

Transposing the conditional: involves converting one kind of conditional probability into a different kind (in mathematical notation, switching round the variables on either side of the ***conditioning bar***). ***Bayes Theorem*** is a formula for effecting this transposition legitimately, by allowing conditional probabilities to be updated in the light of new information. A common reasoning fallacy involves transposing the conditional *illegitimately*. When perpetrated with 'I' (innocence of the defendant) and 'E' (evidence), confusing p(E|I) and p(I|E), it is often described as ***the prosecutor's fallacy***, although the fallacy is by no means confined to prosecutors. A small value for p(E|I) (as in the ***random match probability*** for a DNA profile) does not necessarily mean a small value for p(I|E), the probability of innocence in light of the evidence. A small probability of finding the evidence on an innocent person does not necessarily mean a small probability of innocence for a person on whom the evidence is found. A particularly widespread variant of illegitimately transposing the conditional is ***source probability error***.

Trial: in a statistical context, this is the process by which data are collected in order to investigate some phenomenon thought to be evidenced by those data. For example, a statistical trial might involve repeated tosses of a coin or spins of a roulette wheel. Or a clinical trial could be the process by which the responses of patients to particular drugs are evaluated in order to assess the efficacy of the drug in treating a disease.

Appendix B – Technical Elucidation and Illustrations

Sample Size and Percentages

Sample size is important when considering the precision of estimates. Consider an experimental trial like the example given in §2.7. The sample comprised 1,000 spins of a standard roulette wheel. In percentage terms, the difference between the expected and observed frequencies of the ball landing in the no.1 slot was calculated to be 0.8%; the difference in the absolute frequencies was 35 (observed) to 27 (expected) no.1 slots. Trials comprising 10,000 spins or only 100 spins, however, would be expected to produce, respectively, more or less reliable estimates. As a rule of thumb, the precision of an estimate is related to the square root of the sample size; in order to double the precision of an estimate it is necessary to quadruple the sample size.

Consider another illustration based on coin-tossing. Thirteen heads in twenty tosses of a fair coin (65% heads) is not unusual; using standard probabilistic calculations thirteen or more heads would be expected to occur once in every seven or eight sets of 20 tosses of a fair coin. However, 130 heads in 200 tosses of a fair coin (also 65% heads) *would* be unusual – 130 or more heads would be expected about once in every 550 sets of 200 tosses of a fair coin..

The Multiplication (Product) Rule for Probability[24]

The multiplication rule for probability concerns the conjunction of events. It is best introduced through an artificial example. Consider an urn containing black and white balls in proportions b and w, respectively, where proportions are taken to be numbers between 0 and 1, and b and w are such that $b + w = 1$. The exact number of balls of each colour is not important. In addition to the colour of the balls, assume each ball is either spotted or plain with proportions s and p, and where $s + p = 1$. There are then four types of ball: 'black, spotted', 'black, plain', 'white, spotted' and 'white, plain', denoted c, e, d and f, respectively, such that $c + d + e + f = 1$; $c + d = s$; $e + f = p$; $c + e = b$; and $d + f = w$. These results are conveniently displayed in Table B1.

[24] This section draws on Lindley (1991).

Table B1: Proportions of black, white, spotted and plain balls in an urn

	Black	White	Total
Spotted	c	d	s
Plain	e	f	p
Total	b	w	1

The proportions of spotted and plain balls (*s* and *p*) are given in the final column, labelled 'Total'. The proportions of the black and white balls (*b* and *w*) are given in the final row, also labelled 'Total'.

Let *K* denote the composition of the urn. Let *B* be the event that a ball drawn at random is black and *S* be the event that a ball drawn at random is spotted. Thus, the event that a ball drawn at random is black and spotted is denoted '*B* and *S*'. For conjunctions, the 'and' is often dropped. In this example '*B* and *S*' would be written as *BS*. Proportions can easily be translated into probabilities, since they obey the same rules of logic. Thus, the probability that a ball drawn at random is black, given the composition *K* of the urn, is *b*. Similarly, the probability a ball drawn at random is spotted, given the composition of the urn, is *s*. The probability a ball drawn at random is spotted and black is *c*.

A new idea is now introduced. Suppose someone else had withdrawn a ball at random and announced, truthfully, that it was black. What is the probability that this black ball is also spotted? It is equivalent to the proportion of spotted balls which are also black, which from Table B1 is *c/b*, spotted over black.

Consider the trivial result that

$$c = b \times (c/b).$$

In words, the proportion *c* of balls that are both black and spotted is the proportion *b*, balls that are black, multiplied by the proportion of spotted balls amongst the black balls (*c* out of *b*, or *c/b*).

The equivalent result for probabilities is

$$\mathrm{p}(B \text{ and } S) = \mathrm{p}(B) \times \mathrm{p}(S \mid B).$$

Section 2.35 gives an example of this result applied to the drawing of Aces without replacement from a pack of playing cards. Event B is the drawing of an Ace in the first draw, event S is the drawing of an Ace in the second draw. The left-hand-side of the equation is the drawing of two Aces, which was shown by direct enumeration to have a probability of 1/221. For the right-hand-side, $p(B) = 1/13$ and $p(S \mid B)$ is the probability of drawing an Ace as the second card given that an Ace has been drawn as the first card, which has been shown to be 1/17. The product of 1/13 and 1/17 is 1/221, which is equal to the value on the left-hand-side.

Conditional Probabilities for Dependent Events – A Counter-intuitive Result

One might anticipate that the conditional probability of two dependent events would always be smaller than the probability of the first event taken in isolation. For example, the probability of drawing an Ace from a normal playing deck is 4/52 = 1/13, whereas the probability of drawing an Ace after an Ace has already been drawn without replacement is 3/51 = 1/17. The probability of drawing an Ace after *two* Aces have already been drawn without replacement is even smaller, 2/50 = 1/25.

However, in some cases the probability of an event conditional on another event is actually *greater than* the unconditional probability of the event. Imagine that the frequency of baldness in the general population is 10%. The probability that a person selected at random is bald is therefore 0.10. But notice how these probabilities change if we condition the probability of baldness on gender. Now we would intuitively expect the frequency of baldness *conditioned on being male* to increase, say to 20%; and the frequency of baldness *conditioned on being female* to decrease, say to (almost) 0%. Conditioned on gender, the probability that a person selected at random who is male is also bald is 0.20. And the probability that a person selected at random who is female is also bald is nearly zero. So the frequency of baldness conditioned on gender may be *greater or less* than the unconditional population frequency of baldness.

This result is obtained only for *dependent events*, as where maleness also predicts baldness. If one were to assume independence of baldness and gender, the probability that a person selected at random from the population is bald would remain 0.10 as before, regardless of whether that probability were conditioned on the person's being male, or female, or of unknown gender.

For dependent events only, a conditioning event (gender in the example) may cause the probability of the original event (baldness) to increase or decrease, depending on the nature of the conditioning event.

Interrogating Base Rates

Statistical data, such as those adduced in criminal proceedings as base rates (see §§2.20-2.22, above), need to be interpreted with care. A statistic expressed as a percentage or ***relative frequency*** may be entirely valid, in a formal sense, and yet still potentially seriously misleading. Kaye and Freedman (2000), in their contribution to the US Federal Judicial Center's *Reference Manual on Scientific Evidence*, identify a number of pertinent questions that one might ask when interrogating base rates:

1. *Have appropriate benchmarks been provided?*

   Selective presentation of numerical information can be misleading. Kaye and Freedman (2000) cite a television commercial for a mutual fund trade association which boasted that a $10,000 investment in a mutual trade fund made in 1950 would have been worth $113,500 by the end of 1972. However, according to the *Wall Street Journal*, that same $10,000 investment would have grown to $151,427 if it had been spread over all the stocks comprising the New York Stock Exchange Composite Index.

2. *Have data collection procedures changed?*

   One of the more obvious pitfalls in comparing data time series is that the protocols for data collection may have changed over time. For example, apparent sharp rises or falls in social data, such as morbidity or crime rates, may be mere artefacts of changes in data reporting or recording practices with absolutely no bearing on the underlying social reality.

3. *Are data classifications appropriate?*

   Data can be classified and organised in different ways. One must therefore be alive to the possibility that a particular classification has been selected quite deliberately to support a particular argument or to a highlight a favourable comparison – and by implication to downplay unfavourable arguments or comparisons. Gastwirth (1988b) cites the following example from the USA.

In 1980, tobacco company M sought an injunction to stop the makers of T low-tar tobacco from running advertisements claiming that participants in a national taste test preferred T to other brands. The plaintiffs objected that the advertising claims that T was a "national test winner" and "beats" other brands were false and misleading. In reply, the defendant invoked the data summarised in Table B2 as evidence.

Table B2: The preferences of participants in a national taste test for the comparison of T and M tobacco.

	T much better than M	T somewhat better than M	T about the same as M	T somewhat worse than M	T much worse than M
Number	45	73	77	93	36
Percentage	14	22	24	29	11

According to these data, more survey respondents judged T much better than M (14%) than those finding T much worse than M (11%). Also, 60% regarded T as better or the same as M (i.e. including the 24% who expressed no preference either way). But another way of interpreting these data is to note that 40% who expressed a clear preference actually preferred M to T, whilst only 36% actively preferred T to M. The court ruled in favour of the plaintiffs.

4. *How big is the base of a percentage?*

When the base is small, actual numbers may be more informative than percentages. For example, an increase form 10 to 20 and an increase from 1 million to 2 million are both 100% increases. To say that something has increased "by 100 per cent" always sounds impressive, but whether it is or not depends, amongst other things, on the numbers behind the percentage. (Also recall the coin-tossing examples of 13 heads in 20 tosses and 130 heads in 200 tosses, discussed in the first section of this Appendix.)

5. *Which comparisons are made?*

Comparisons are always made relative to some base-line, so that the choice of base-line (where eligible alternatives are available) may be a crucial factor in interpreting the meaning of any statistic. Suppose that a University reports that the proportion of first class degrees awarded in humanities subjects has increased by 30% on the previous year. All well and good. But is the previous year an appropriate base-line? What if the previous year was a markedly fallow year for first class degrees in the humanities, so that a 30% increase merely restores the level of firsts to what it was two years ago? Conversely, there may have been a big increase in firsts in the previous year as well, perhaps suggesting a worrying erosion in academic standards rather than an impressive improvement in student performance. In this and many other similar scenarios, choice of base-line has a major bearing on the meaning – and probative value – of statistical information.

Illegitimately transposing the conditional – case illustrations

There are numerous reported cases involving illegitimate transpositions of the conditional ("***the prosecutor's fallacy***"). This is how it occurred in *Deen*[25] in relation to a DNA profile with a frequency of 1 in 3 million in the relevant population:

> *Prosecuting counsel*: So the likelihood of this being any other man but Andrew Deen is one in 3 million?
>
> *Expert*: In 3 million, yes.
>
> *Prosecuting counsel*: You are a scientist... doing this research. At the end of this appeal a jury are going to be asked whether they are sure that it is Andrew Deen who committed this particular rape in relation to Miss W. On the figure which you have established according to your research, the possibility of it being anybody else being one in 3 million what is your conclusion?
>
> *Expert*: My conclusion is that the semen originated from Andrew Deen.
>
> *Prosecuting counsel*: Are you sure of that?
>
> *Expert*: Yes.

[25] *R* v *Deen*, CA, *The Times*, 10 January 1994.

The fallacy is perpetrated when the expert is induced to agree that the likelihood (probability) of the criminal being someone other than Andrew Deen, given the evidence of the DNA match, is one in three million. (This error was further compounded by the unwarranted source-level conclusion that Deen was the source of the stain, i.e. ***source probability error***.)

The relative frequency of the DNA profile in the relevant population was 1 in 3 million, meaning that one person in every 3 million selected at random from this population would be expected to have a matching profile. This is patently not the probability that a person with a matching profile is innocent, as the quoted exchange between the expert and prosecuting counsel clearly implies. The conditional has been transposed illegitimately. One cannot calculate the probability of guilt or innocence of a particular person without knowing the number of people in the relevant suspect population. If the suspect population comprised, say, 6 million individuals, one would expect two matching profiles amongst the innocent people. Add this to the offender (whose probability of matching can be taken to be 1) and the expected number of people with the profile is 3, giving a probability of guilt for a person with the profile – $p(G|E) = 1/3$.

An expert witness called by the prosecution also illegitimately transposed the conditional in *Doheny and Adams*, as recounted by the Court of Appeal:[26]

> "A. 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.
> Q. The likelihood of it being anybody other than Alan Doheny?
> A. Is about 1 in 40 million.
> Q. You deal habitually with these things, the jury have to say, of course, on the evidence, whether they are satisfied beyond doubt that it is he. You have done the analysis, are you sure that it is he?
> A. Yes."
> The question, in leading form, and the numerical answer given to it constituted a classic example of the 'prosecutor's fallacy'. The third question was one for the jury, not for the witness. The witness gave an affirmative answer to it. It is not clear to what evidence, if any, other than the DNA evidence, he had regard when giving that answer. For the reasons that we gave in our introduction to this Judgment, this series of questions and answers was inappropriate and potentially misleading.

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

A third illustration comes from *Gordon*,[27] where the relative frequencies of the DNA profiles in question were calculated to be 1 in ten-and-a-half million and 1 in just over seventeen million. An expert witness testified that 'she was sure of the match between the semen samples and the appellant's blood'.[28] This is ***source probability error***, since even the extreme unlikelihood of a random match does not permit the expert to infer a definitive source. Fundamentally, to confuse the probability that a DNA profile derived from a crime scene will match an innocent person's profile (the ***random match probability***) with the probability that a person with a matching profile is innocent, as the expert appears to have done in *Gordon*, is to commit the fallacy of illegitimately transposing the conditional.

Calculating the probability of "another match"

As we explained in §, the probability of finding "another match" should not to be confused with the ***random match probability***. Here is the more technical explanation.

Consider a characteristic which is prevalent in only 1 in a thousand, 1/1,000, people (e.g. a height greater than a certain designated value, such as two metres). It is sometimes claimed that the significance of evidence of this characteristic can be expressed in terms of the number of people who would have to be counted before there is another (random) match, being the reciprocal of the frequency (1,000, in this example); i.e. "1,000 people would need to be observed before someone else of that height would be encountered". Yet this is an intuitively obvious fallacy, since the very next person observed could be that height or taller.

This result can be demonstrated formulaically. It has been established that the probability that a person is no taller than two metres is 999/1,000. If n independent (unrelated) people are observed, we also know by repeated use of ***the product rule for independent events*** that the probability that none is taller than two metres is $(999/1000)^n$ (the probability is 999/1000 on each selection, and we make n independent selections). The complementary event is that at least one person is taller than two metres in height, i.e. $1 - (999/1000)^n$. For it to be more likely than not that at least one person is taller than two meters, 1 -

[27] *R* v *Gordon* [1995] 1 Cr App R 290, CA.

[28] ibid. 293.

$(999/1000)^n$ must be greater than 0.5. In fact the formula 1 - $(999/1000)^n$ equals 0.5 when $n = 693$, so it is more likely than not that at least one person will be taller than two metres after selecting 694 people – *not* after 1,000 selections. If 1,000 people were indeed observed, the probability that at least one of them would be over two metres in height is 0.632. In order to raise the probability of at least one other person of at least that height to 0.9 one would need to look at 2,307 people, which is the value of n where 1 - $(999/1000)^n$ = 0.9.

General Principles for the Presentation of Scientific Evidence

Various attempts have been made over the years to formulate general principles to guide the presentation and interpretation of scientific and other expert evidence in criminal proceedings. Here, for ease of reference, we summarise two significant sources of normative guidance.

First, Part 33 (Expert Evidence) of the Criminal Procedure Rules 2010 includes the following requirements:

Rule 33.2 - Expert's duty to the court

(1) An expert must help the court… by giving objective, unbiased opinion on matters within his expertise.

(2) This duty overrides any obligation to the person from whom he receives instructions or by whom he is paid.

(3) This duty includes an obligation to inform all parties and the court if the expert's opinion changes from that contained in a report served as evidence or given in a statement.

Rule 33.3 - Content of expert's report

(1) An expert's report must—

- (a) give details of the expert's qualifications, relevant experience and accreditation;
- (b) give details of any literature or other information which the expert has relied on in making the report;
- (c) contain a statement setting out the substance of all facts given to the expert which are material to the opinions expressed in the report, or upon which those opinions are based;
- (d) make clear which of the facts stated in the report are within the expert's own knowledge;
- (e) say who carried out any examination, measurement, test or experiment which the expert has used for the report and—

(i) give the qualifications, relevant experience and accreditation of that person,

(ii) say whether or not the examination, measurement, test or experiment was carried out under the expert's supervision, and

(iii) summarise the findings on which the expert relies;

(f) where there is a range of opinion on the matters dealt with in the report—

(i) summarise the range of opinion, and

(ii) give reasons for his own opinion;

(g) if the expert is not able to give his opinion without qualification, state the qualification;

(h) contain a summary of the conclusions reached;

(i) contain a statement that the expert understands his duty to the court, and has complied and will continue to comply with that duty; and

(j) contain the same declaration of truth as a witness statement.

These criteria for expert report writing may be regarded *mutatis mutandis* as general expectations of scientific evidence adduced in legal proceedings in any form, including live oral testimony. The Court of Appeal has reiterated the vital importance of full compliance with CrimPR 2010 Rule 33 on many occasions.

Further normative guidance might be found in the following list of criteria and associated principles, which have been advanced by the Association of Forensic Science Providers:[29]

- *Balance*: The expert should address at least one pair of propositions.
- *Logic*: The expert will address the probability of the evidence given the proposition and relevant background information and not the probability of the proposition given the evidence and background information.
- *Robustness*: The expert will provide factual and opinion evidence that is capable of scrutiny by other experts and cross-examination. Expert evidence will be based on sound knowledge of the evidence type(s) and use verified databases, wherever possible.

[29] The Association of Forensic Science Providers aims to "represent the common interests of the providers of independent forensic science within the UK and Ireland with regard to the maintenance and development of quality and best practice in forensic science and expert witness in support of the Justice System, from scene to court, irrespective of the commercial pressures associated with the competitive forensic marketplace": see Brown and Willis (2009); Association of Forensic Science Providers (2009).

- *Transparency*: The expert will be able to demonstrate how inferential conclusions were produced: propositions addressed, examination results, background information, data used and their provenance.

These desiderata for expert evidence encapsulate several of the points stressed in this Report. The first principle expresses the idea that it is not sufficient to consider the value of evidence – even strongly incriminating evidence – in the abstract. Evidential value is a function of two competing propositions, the likelihood of the evidence on the assumption that the prosecution's proposition is true and the likelihood of the evidence on the assumption that the prosecution's proposition is false. The second principle reiterates the elementary injunction against illegitimately transposing the conditional. As a general rule, forensic scientists and other expert witnesses should be assessing the probability of the evidence, rather than commenting on the probability of contested facts (much less the ultimate issue of guilt or innocence). Robustness is concerned with scientific methodology, which must be valid and able to withstand appropriately searching scrutiny. The knowledge of the expert must be sound. Laboratory equipment must be in good working order, properly calibrated. Operational protocols should be validated with known error rates. Databases will have been verified or accredited as much as possible. Finally, the principle of transparency states that all of the assumptions, data, instrumentation and methods relied on in producing the evidence must stated explicitly or at least open to examination and verification by the court.

Appendix C – Select Case Law Precedents and Further Illustrations

1. English and UK Law

***Pringle* v *R,* Appeal No. 17 of 2002, PC(Jam) –** illustrates a range of difficulties with the probabilistic interpretation of DNA evidence, inc: unwarranted assumptions of independence; "prosecutor's fallacy" (illegitimately transposing the conditional) at trial; apparent misunderstanding of statistical frequencies on appeal.

***R* v *Adams (No 2)* [1998] 1 Cr App R 377, CA –** juries employ common sense reasoning in reaching their verdicts in criminal cases, and should not be encouraged by expert witnesses to employ mathematical formulae, such as Bayes' Theorem, to augment – or more likely confuse – their ordinary reasoning processes (reiterating ***R* v *Adams* [1996] 2 Cr App R 467, CA**).

***R* v *Atkins* [2010] 1 Cr App R 8, [2009] EWCA Crim 1876 –** expert witness in "facial mapping" permitted to express conclusions about the strength of his evidence in terms of a (non-mathematical or statistical) six-point scale utilising expressions such "lends support", "lends strong support", etc.

***R* v *Benn and Benn* [2004] EWCA Crim 2100 –** judicial consideration of the adequacy of databases (here, in relation to patterns of cocaine contamination on banknotes).

***R* v *Bilal* [2005] EWCA Crim 1555** – illustration of source probability error in relation to handwriting samples.

***R* v *Clark* [2003] EWCA Crim 1020 –** unwarranted assumption of independence, leading to inappropriate use of the product rule for independent events to calculate a fallacious probability of multiple sudden infant deaths (SIDS) in the same family.

***R* v *Dallagher* [2003] 1 Cr App R 12, [2002] EWCA Crim 1903** –.expert was permitted to testify that D was very likely to be the donor of an earprint at the scene of the crime, on the explicit assumption that earprints are uniquely identifying (notwithstanding the paucity of the research base justifying this assumption). *Semble* there is no source probability error if the probability of an innocent match is zero; though it is difficult to see how this assumption can ever be valid in the real world.

***R* v *Deen, The Times,* 10 January 1994 (CA, 21 December 1993)** – early example of "the prosecutor's fallacy" (illegitimately transposing the conditional) leading to conviction being quashed on appeal.

***R* v *Doheny and Adams* [1997] 1 Cr App R 369, [1996] EWCA Crim 728** – general discussion of the "prosecutor's fallacy" (illegitimately transposing the conditional). DNA experts should testify to the "random occurrence ratio" (random match probability) rather than expressing any inferential conclusion about the donor of suspect DNA.

***R* v *George (Barry)* [2007] EWCA Crim 2722** – application of basic principles of relevance and probative value to scientific evidence. The court heard evidence that the scientific findings were equally likely to be obtained if Mr George was or was not the person who had shot the victim, Jill Dando. If, as other evidence suggested, it was just as likely that a single particle of firearms discharge residue (FDR) came from some extraneous source as it was that it came from a gun fired by the appellant, it was misleading to tell the jury that innocent contamination was "most unlikely" (with the apparent implication that the FDR evidence must therefore be materially incriminating).

***R* v *Gordon* [1995] 1 Cr App R 290, CA** – early illustration indicating some of the practical problems that may arise in relation to DNA evidence, inc: contested criteria for declaring a "match" between samples; and adequacy of choice of reference class (population database) and its bearing on the random match probability.

***R* v *Gray (Kelly)* [2005] EWCA Crim 3564** – illustration of DNA expert inadvertently being tempted into source probability error by questions put in cross-examination. These slips were not regarded as affecting the safety of the conviction, where the value of the evidence has previously been correctly stated by the expert.

***R* v *Gray (Paul Edward)* [2003] EWCA Crim 1001** - CA cast doubt on an expert's ability to make positive identifications using facial mapping techniques in the absence of reliable databases of facial characteristics. However, these remarks were distinguished in ***R* v *Atkins* [2010] 1 Cr App R 8, [2009] EWCA Crim 1876.**

***R* v *Reed and Reed; R* v *Garmson* [2010] 1 Cr App R 23; [2009] EWCA Crim 2698 –** provided that the basis for the opinion is clearly set out (and that this is properly reflected in the trial judge's direction to the jury), an expert may present inferential conclusions about the likely provenance of biological material from which a DNA profile was extracted. Such testimony may incorporate unquantified probabilities of transfer and persistence, but must not advance speculative activity level propositions lacking any truly scientific basis.

***R* v *Robb* (1991) 93 Cr App R 161, CA –** expert witness is permitted to form opinion on basis of unquantified experience expressing minority view in the field; affirmed in ***R* v *Flynn and St John* [2008] 2 Cr App R 20, [2008] EWCA Crim 970**.

***R* v *Shillibier* [2006] EWCA Crim 793 –** example of source probability error in making comparisons between soil samples.

***R* v *Stockwell* (1993) 97 Cr App R 260, CA –** continued existence of a strict "ultimate issue rule" doubted; reiterated in ***R* v *Atkins* [2009] EWCA Crim 1876**.

***R* v *T* [2010] EWCA Crim 2439 –** the "Bayesian approach" to evaluating evidence, employing likelihood ratios, should be confined to types of evidence (such as DNA profiling) for which there exist reliable databases. In the current state of knowledge, expertise in footwear mark comparison does not meet this standard, and consequently should be limited to the expression of non-probabilistic evaluative opinions.

***R* v *Weller* [2010] EWCA Crim 1085 –** expert witness permitted to express conclusions about source, transfer, and persistence of genetic material based partly on experience and unpublished research.

2. Foreign and Comparative Sources

***Hughes* v *State*, 735 So 2d 238 (1999), Supreme Court of Mississippi –** explicit recognition and discussion of numerical conversion error.

***People* v *Collins*, 68 Cal 2d 319, 66 Cal Rptr 497 (1968), Supreme Court of California** – classic illustration of the misuses of forensic probability, including speculative relative frequency values with no evidential basis and unsubstantiated assumptions of independence when utilising the product rule for independent events.

***R* v *Montella* [1992] 1 NZLR 63, High Court** – a first instance ruling on admissibility, illustrating the use of a likelihood ratio to express the probative value of expert DNA evidence: "It is said that the likelihood of obtaining such DNA profiling results is at least 12,400 times greater if the semen stain originated from the accused than from another individual".

***State* v *Bloom*, 516 N W 2d 159 (1994), Supreme Court of Minnesota** – clear exposition of source probability error and other common mistakes in probabilistic reasoning, and consideration of how probabilistic evidence might best be presented to juries.

***Smith* v *Rapid Transit*, 317 Mass 469, 58 N E 2d 754 (1945), Supreme Judicial Court of Massachusetts** – this very short judgment, upholding a directed verdict for the defendant in a negligence action, inspired the much discussed "Blue Bus" hypothetical and related problems associated with proof by "naked statistical evidence": see, e.g., Redmayne (2008).

***US* v *Shonubi,* 895 F Supp 460 (EDNY, 4 Aug 1995) ["*Shonubi* III"]** – Judge Weinstein reviewed the general principles of forensic statistics.

***Wike* v *State*, 596 So 2d 1020 (1992), Supreme Court of Florida** – an illustration of source probability error. Whereas other physical trace evidence adduced by the prosecution is correctly summarized as being (merely) "consistent with" the accused or the victim being its donor, a DNA profile of a blood sample is erroneously described as "positively coming from" the victim.

***Williams* v *State*, 251 Ga 749, 312 S E 2d 40 (1983), Supreme Court of Georgia –** Justice Smith, dissenting, makes a number of pertinent points challenging the adequacy of the prosecution's carpet fibre evidence, which was expressed to the jury in terms of a compound relative frequency of one in forty million. Smith J. objects that the individual relative frequencies which went into this calculation were mere surmises which were insufficiently proved by admissible evidence.

Appendix D – Select Bibliography

Aitken, C.G.G. and Taroni, F. (2004) *Statistics and the Evaluation of Evidence for Forensic Scientists*. Chichester: Wiley.

- (2008) 'Fundamentals of Statistical Evidence – A Primer for Legal Professionals' 12 *International Journal of Evidence & Proof* 181.

Allen, R.J. (1991) 'The Nature of Juridical Proof' (1991) 13 *Cardozo LR* 373.

Allen, R.J. and Pardo, M. (2007) 'The Problematic Value of Mathematical Models of Evidence' 36 *Journal of Legal Studies* 107.

Allen, R.J. and Redmayne, M. (eds.) (1997) *Special Issue on Bayesianism and Juridical Proof* 1(6) *International Journal of Evidence & Proof* 253.

Allen, R.J. and Roberts, P. (eds.) (2007), *Special Issue on the Reference Class Problem* 11(4) *International Journal of Evidence & Proof* 243.

Association of Forensic Science Providers (2009) 'Standards for the Formulation of Evaluative Forensic Science Expert Opinion' 49 *Science and Justice* 161.

Balding, D.J. (2005) *Weight-of-Evidence for Forensic DNA Profiles*. Chichester: Wiley.

Balding, D.J. and Donnelly, P. (1994) 'The Prosecutor's Fallacy and DNA Evidence' *Criminal Law Review* 711.

Brown, S. and Willis, S. (2009) 'Complexity in Forensic Science' 1 *Forensic Science Policy and Management* 192.

Buckleton J.S. (2004) 'Population Genetic Models' in J.S. Buckleton, C.M. Triggs and S.J. Walsh (eds.) *DNA Evidence*. Boca Raton, Florida: CRC Press.

Callen, C.R. (1982) 'Notes on a Grand Illusion: Some Limits on the Use of Bayesian Theory in Evidence Law' 57 *Indiana Law Journal* 1.

- (1991) 'Adjudication and the Appearance of Statistical Evidence' 65 *Tulane Law Review* 457.

Champod C., Evett I.W. and Jackson, G. (2004) 'Establishing the Most Appropriate Databases for Addressing Source Level Propositions' 44 *Science and Justice* 153.

Coleman, R.F. and Walls, H.J. (1974) 'The Evaluation of Scientific Evidence' *Criminal Law Review* 276.

Cook, R., Evett, I.W., Jackson, G., Jones, P.J. and Lambert, J.A. (1998a) 'A Model for Case Assessment and Interpretation' 38 *Science & Justice* 151.

- (1998b) ‘A Hierarchy of Propositions: Deciding which Level to Address in Casework’ 38 *Science & Justice* 231.
- (1999) ‘Case Pre-assessment and Review of a Two-way Transfer Case’ 39 *Science & Justice* 103.

Dawid, A. P. (2005) ‘Probability and Proof’, on-line Appendix I to T.J. Anderson, D.A. Schum and W.L. Twining, *Analysis of Evidence: Second Edition*. Cambridge: CUP. http://tinyurl.com/7g3bd (accessed 19 October 2010).

DeGroot, M. H., Fienberg, S.E. and Kadane, J.B. (eds.) (1994) *Statistics and the Law*. New York: Wiley.

Diamond, S.S. (2000) ‘Reference Guide on Survey Research’ in *Reference Manual on Scientific Evidence*, 2nd edn. Federal Judicial Center: Washington, DC.

Eggleston, R. (1983) *Evidence Proof and Probability,* 2nd edn. London: Butterworths.

Evett, I.W., Foreman, L.A., Jackson, G. and Lambert, J.A. (2000) ‘DNA Profiling: A Discussion of Issues Relating to the Reporting of Very Small Match Probabilities’ *Criminal Law Review* 341.

Evett, I.W., Jackson, G., Lambert, J.A. and McCrossan, S. (2000) ‘The Impact of the Principles of Evidence Interpretation and the Structure and Content of Statements’ 40 *Science & Justice* 233.

Evett, I.W. and Weir, B.S. (1998) *Interpreting DNA Evidence.* Sunderland, Mass.: Sinauer Associates Inc.

Fienberg, S. E. (ed.) (1989) *The Evolving Role of Statistical Assessments as Evidence in the Courts*. New York: Springer.

Finkelstein, M. (2009) *Basic Concepts of Probability and Statistics in the Law*. New York: Springer.

Finkelstein, M.O. and Levin, B. (2001) *Statistics for Lawyers*, 2nd edn. New York: Springer.

Fleming, P., Blair, P., Bacon, C., Berry, J. (2000) *Sudden Unexpected Deaths in Infancy.* London: HMSO.

Friedman, R.D. (1996) ‘Assessing Evidence’ 94 *Michigan Law Review* 1810.

Gastwirth, J.L. (1988a) *Statistical Reasoning in Law and Public Policy, vol 1: Statistical Concepts and Issues of Fairness*. Boston, Mass.: Academic Press.

- (1988b) *Statistical Reasoning in Law and Public Policy, vol 2: Tort Law, Evidence and Health*. Boston, Mass.: Academic Press.
- (ed.) (2000) *Statistical Science in the Courtroom*. New York: Springer.

Hodgson, D. (1995) ‘Probability: The Logic of the Law – A Response’ 15 *Oxford Journal of Legal Studies* 51.

Holden, C. (1997) ‘DNA Fingerprinting Comes of Age’ 278 *Science* 1407.

Jackson, G. (2009) ‘Understanding Forensic Science Opinions’ in J. Fraser and R. Williams (eds.), *Handbook of Forensic Science*. Cullompton, Devon: Willan Publishing.

Jackson, G., Jones, S., Booth, G., Champod, C. and Evett, I.W. (2006) ‘The Nature of Forensic Science Opinion - A Possible Framework to Guide Thinking and Practice in Investigations and in Court Proceedings’ 46 *Science and Justice* 33.

Kadane, J.B. (2008) *Statistics in the Law: A Practitioner’s Guide, Cases, and Materials*. New York: OUP.

Kaye, D.H. (1979) ‘The Laws of Probability and the Law of the Land’ 47 *University of Chicago Law Review* 34.

- (1984) ‘Thinking Like a Statistician: The Report of the American Statistical Association Committee on Training in Statistics in Selected Professions’ 34 *Journal of Legal Education* 97.

- (1993) ‘DNA Evidence: Probability, Population Genetics and the Courts’ 7 *Harvard Journal of Law and Technology* 101.

Kaye, D.H. and Freedman, D.A. (2000) ‘Reference Guide on Statistics’ in *Reference Manual on Scientific Evidence*, 2nd edn. Federal Judicial Center: Washington, DC.

Koehler, J.J. (1993) ‘Error and Exaggeration in the Presentation of DNA Evidence at Trial’ 34 *Jurimetrics Journal* 21.

- (2001) ‘The Psychology of Numbers in the Courtroom: How to Make DNA-Match Statistics Seem Impressive or Insufficient’ 74 *Southern California Law Review* 1275.

Koehler, J.J. and Shaviro, D.N. (1990) ‘Veridical Verdicts: Increasing Verdict Accuracy Through the Use of Overtly Probabilistic Evidence and Methods’ 75 *Cornell Law Review* 247.

Lempert, R.O. (1977) ‘Analyzing Relevance’ 75 *Michigan Law Review* 1021.

- (1991) ‘Some Caveats Concerning DNA as Criminal Identification Evidence: with Thanks to the Reverend Bayes’ 13 *Cardozo Law Review* 303.

- (1993) ‘The Suspect Population and DNA Identification’ 34 *Jurimetrics Journal* 1.

Lindley, D.V. (1991) ‘Probability’ in C.G.G Aitken and D.A. Stoney (eds.) *The Use of Statistics in Forensic Science*. Chichester: Ellis Horwood.

Redmayne, M. (1997) ‘Presenting Probabilities in Court: The DNA Experience’ 1 *International Journal of Evidence & Proof* 187.

- (1998) ‘Bayesianism and Proof’ in M. Freeman and H. Reece (eds.), *Science in Court*. Aldershot: Ashgate.
- (2001) *Expert Evidence and Criminal Justice*. Oxford: OUP.
- (2008) ‘Exploring the Proof Paradoxes’ 14 *Legal Theory* 281.

Roberts, P. (2009) ‘The Science of Proof: Forensic Science Evidence in English Criminal Trials’ in J. Fraser and R. Williams (eds.), *Handbook of Forensic Science*. Cullompton, Devon: Willan Publishing.

Roberts, P. and Zuckerman, A. (2010) *Criminal Evidence*, 2nd edn. Oxford: OUP.

Robertson, B. and Vignaux, G.A. (1992) ‘Unhelpful Evidence in Paternity Cases’ 9 *New Zealand Law Journal* 315.

- (1993) ‘Probability - The Logic of the Law’ 13 *Oxford Journal of Legal Studies* 457.
- (1995) *Interpreting Evidence: Evaluating Forensic Science in the Courtroom.* Chichester: Wiley.

Saks, M.J. and Koehler, J.J. (2005) ‘The Coming Paradigm Shift in Forensic Identification Science’ 309 *Science* 892.

Schum, D.A. (1994) *The Evidential Foundations of Probabilistic Reasoning.* Evanston, Illinois: Northwestern UP.

Stoney, D.A. (1991) ‘What Made Us Think We Could Individualise Using Statistics?’ 31 *Journal of the Forensic Science Society* 197.

Thompson, W.C. and Schumann, E.L. (1987) ‘Interpretation of Statistical Evidence in Criminal Trials: The Prosecutor’s Fallacy and the Defence Attorney’s Fallacy’ 11 *Law and Human Behaviour* 167.

Thompson, W.C, Taroni, F., and Aitken, C.G.G. (2003) ‘How the Probability of a False Positive Affects the Value of DNA Evidence’ 48 *Journal of Forensic Sciences* 47.

Tribe, L.H. (1971) ‘Trial by Mathematics: Precision and Ritual in the Legal Process’ 84 *Harvard Law Review* 1329.

Usher, M.A. and Stapleton, R.R. (1979) ‘An Unusual Forensic Application of Blood Group Studies’ 19 *Medicine, Science and the Law* 165.

Zeisel, H. and Kaye, D. (1997) *Prove it with Figures: Empirical Methods in Law and Litigation*. New York: Springer.