

---

# Communicating and Interpreting Statistical Evidence in the Administration of Criminal Justice

---

## 2. Assessing the probative value of DNA evidence

---

Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses

**Roberto Puch-Solis**  
**Paul Roberts**  
**Susan Pope**  
**Colin Aitken**



PRACTITIONER GUIDE NO 2

## **Assessing the Probative Value of DNA Evidence**

*Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses*

By

**Roberto Puch-Solis**, Forensic Statistician, Forensic Science Service

**Paul Roberts**, Professor of Criminal Jurisprudence, University of Nottingham

**Susan Pope**, Principal DNA Scientist, Forensic Science Service

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

Prepared under the auspices of the  
Royal Statistical Society's *Working Group on Statistics and the Law*  
(Chairman: Colin Aitken)

## Contents

0.	Introduction to Communicating and Interpreting Statistical Evidence in the Administration of Criminal Justice	2
	Membership of the RSS Working Group on Statistics and the Law	11
1.	DNA Evidence in Criminal Proceedings	12
	<i>Potted history</i> .....	12
	<i>The probabilistic foundations of DNA evidence</i> .....	13
	<i>Putting DNA in its probative place</i> .....	15
	<i>DNA and the law of evidence</i> .....	18
2.	The DNA Profile	21
	<i>Basic biology</i> .....	21
	<i>Profiling nuclear DNA</i> .....	23
	<i>Obtaining an STR DNA profile</i> .....	24
	<i>Protocols, multiplexes and systems</i> .....	27
3.	DNA Profiles as Evidence in Criminal Proceedings	30
	<i>Source, classification and quality of samples</i> .....	30
	<i>Types of questioned profile</i> .....	31
	<i>Accounting for peak height variation and natural artefacts</i> .....	32
	<i>Propositions in the evaluation of DNA evidence</i> .....	35
4.	Assessing the Probative Value of Single Donor Profiles	39
	<i>The logic of probative value</i> .....	39
	<i>The probative logic of DNA evidence</i> .....	39
	<i>DNA likelihood ratios</i> .....	43
	<i>Allele frequencies</i> .....	47
	<i>Sampling allowance</i> .....	49
	<i>Co-ancestry allowance</i> .....	50
	<i>Blood relatives</i> .....	52
5.	Two Person ('Mixed') Questioned Profiles	57
	<i>Transfer patterns</i> .....	57
	<i>(a) Questioned profile with four balanced peaks for Scenario #1</i> .....	60
	<i>(b) Questioned profile with three peaks for Scenario #1</i> .....	63
	<i>(c) Questioned profile with three unbalanced peaks for Scenario #1</i> .....	67
	<i>(d) Questioned profile with four balanced peaks for Scenario #2</i> .....	69
6.	Low Template DNA (LTDNA)	73
	<i>From LCN to LTDNA</i> .....	73
	<i>Practical benefits of LTDNA profiling</i> .....	74
	<i>Profiling enhancements</i> .....	75
	<i>Enhanced anti-contamination procedures</i> .....	76
	<i>Issues of interpretation</i> .....	77
	<i>Assessing the probative value of LTDNA profiles</i> .....	79
	<i>Example #1: one-person profile with larger heterozygote balance and stutter ratio</i> ....	79
	<i>Example #2: one-person profile with possible allelic dropout</i> .....	82
	<i>Example #3: one-person profile with an additional small peak</i> .....	84
	<i>Assessing the probative value of LTDNA profile evidence</i> .....	85
7.	Presenting DNA Evidence in the Courtroom	88
	<i>The ruling in Doheny and Adams</i> .....	90
	<i>Illegitimately transposing the conditional ('the prosecutor's fallacy')</i> .....	94
	Appendices	
A.	Bibliography	101
B.	The UK DNA Database and Familial Searching	106
C.	Y-STR Profiles	111
D.	Mitochondrial DNA Profiles	114
E.	Glossary	117

## **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 Criminal adjudication in the UK's legal jurisdictions 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 ensure that statistical evidence and probabilities are presented to fact-finders in as clear and comprehensible a fashion as possible. Yet legal professionals' grasp of statistics and probability may in reality 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.<sup>1</sup> 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

---

<sup>1</sup> *R v Atkins* [2010] 1 Cr App R 8, [2009] EWCA Crim 1876; *R v Stockwell* (1993) 97 Cr App R 260, CA; *R v Silverlock* [1894] 2 QB 766, CCR.

even rudimentary – knowledge of 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 their existing levels of knowledge and competence in using statistical methods 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 enough already, but do so only imperfectly ('a little learning is a dang'rous thing'<sup>2</sup>).

0.4. If a broad programme of education for lawyers and other forensic practitioners is needed, what is required 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. Every participant in criminal proceedings should be able to grasp at least enough to perform their respective allotted roles effectively and to discharge their professional responsibilities 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

---

<sup>2</sup> Alexander Pope, *An Essay on Criticism* (1711).

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 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.<sup>3</sup>

0.5 There are doubtless many ways in which these pressing educational needs might be met, possibly through a package of measures and programmes. Of course, design and regulation of professional education are primarily matters to be determined by the relevant professional bodies and regulatory authorities. 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* (Third Edition, 2011) provides a valuable and instructive template.<sup>4</sup> 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

---

<sup>3</sup> *R v B(T)* [2006] 2 Cr App R 3, [2006] EWCA Crim 417, [176]. And see CrimPR 2011, Rule 33.2: 'Expert's duty to the court'.

<sup>4</sup> The recently-revised third edition of the *Reference Manual* is published jointly by the National Academy of Sciences and the Federal Judicial Center, and can be accessed at [www.fjc.gov/public/home.nsf/autoframe?openform&url\\_1=/public/home.nsf/inavgeneral?openpage&url\\_r=/public/home.nsf/pages/1448](http://www.fjc.gov/public/home.nsf/autoframe?openform&url_1=/public/home.nsf/inavgeneral?openpage&url_r=/public/home.nsf/pages/1448).

guidance by producing a themed set of four *Practitioner Guides* on different aspects of statistical evidence and probabilistic reasoning, 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 principally 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 Faculty of Advocates, the Crown Prosecution Service, the National Policing Improvement Agency (NPIA)<sup>5</sup> and the Forensic Science Service,<sup>6</sup> as well as academic lawyers, statisticians and forensic scientists.

#### 0.7 *Using the Four Practitioner Guides – Notes, Caveats and Disclaimers*

The four *Practitioner Guides* are being written over a four-year period, with the final *Guide* scheduled for publication in 2013. They are intended, when completed, to form a coherent package, but each *Guide* is also designed to function as a stand-alone publication addressing a specific topic or set of related issues in detail. Some of the material restates elementary principles and general background that every criminal justice practitioner really ought to know. More specialist sections of the *Guides* might be dipped into for reference as and when occasion demands. 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 statistical or probability-related issues arise during the course of criminal litigation.

---

<sup>5</sup> The NPIA seat on our working group is currently vacant, following the Government's announcement that it intends to phase out NPIA and reallocate its critical functions to other agencies: see House of Commons Home Affairs Committee, *New Landscape of Policing*. Fourteenth Report of Session 2010–12, HC 939 (TSO, 2011).

<sup>6</sup> The Government has announced that the Forensic Science Service will be wound down and cease operations by March 2012: see House of Commons Science and Technology Committee, *The Forensic Science Service*. Seventh Report of Session 2010–12, HC 855 (TSO, 2011).



0.8 *Guide No 1* was published in December 2010 as **Colin Aitken, Paul Roberts and Graham Jackson**, *Fundamentals of Probability and Statistical Evidence in Criminal Proceedings* (RSS, 2010), and is available free to download from the RSS website: [www.rss.org.uk/statsandlaw](http://www.rss.org.uk/statsandlaw).

The first *Guide* provides 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’).

Building on this general introduction, *Guide No 2* explores in the following pages the probabilistic foundations of DNA profiling evidence and considers how to evaluate its probative value in criminal trials. The remaining two *Guides* will give detailed consideration to: (3) networks for structuring evidence; and (4) principles of forensic case assessment and interpretation. Each of these topics has major practical importance, and therefore merits sustained investigation, in its own right. Their systematic exploration will also serve to elucidate the general themes, questions, concepts and issues affecting the communication and interpretation of statistical evidence and probabilistic reasoning in the administration of criminal justice which are addressed across all four *Guides*.

0.9 We should flag up at the outset certain methodological challenges confronting this ambitious undertaking, not least because it is unlikely that we have overcome them all 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 makes sense to pitch the discussion at this generic level. All participants in the process would benefit from improved understanding of other professional groups’ perspectives, assumptions, concerns and objectives. For example, lawyers might adapt and enhance

the ways in which they instruct experts and adduce 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' assumptions and expectations of expert evidence, in particular regarding the extent and implications of its probabilistic underpinnings.

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. Our hope is that judges, lawyers and forensic scientists will be able to extrapolate from the common core and adapt our generic analysis of DNA profiling evidence to the particular demands of their own professional role in criminal proceedings. We have stopped well short of presuming to specify formal criteria of legal admissibility or to formulate concrete guidance that judges might repeat to juries in criminal trials. It is not for us to make detailed recommendations on the law and practice of criminal procedure.

0.10 The following exposition is also generic in a second, related sense. This *Guide* is intended to be useful, and to 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 particular 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 this *Guide's* scope and coverage. Whilst it would defeat our purpose to try to replicate the technical detail of existing DNA literature addressed to specialist scientific audiences, we do aim to provide more in-depth analysis and

discussion than is typically found in the skeleton summaries and ‘check-lists’ currently available to criminal practitioners (see e.g. Forensic Science Service, 2004). Useful as these summaries are, as far as they go, there would be little point in us merely replicating information in a form that is already widely available. We have, instead, tried to strike an appropriate balance between detail, utility and intelligibility. This *Guide* provides just enough information about the basic science of DNA profiling to enable readers to make informed judgements about the probative value of DNA evidence. We concentrate, in particular, on the probabilistic underpinnings of DNA profiling and their evidential implications in criminal adjudication.

We are assuming a non-specialist audience for scientific discussion; albeit an audience comprised of criminal justice professionals with both a strong occupational interest and a professional duty to acquaint themselves with the fundamentals of DNA evidence. In keeping with our broad-brush approach to the law, we have accordingly endeavoured to keep scientific and statistical technicalities to a minimum in the main text. Appendix E, located for ease of reference at the end of this *Guide*, contains a glossary of specialist DNA-related terminology. Three further appendices provide supplementary information regarding the UK national DNA database (NDNAD) and familial searching (Appendix B), Y-STR profiles (Appendix C) and mitochondrial DNA profiles (Appendix D). Appendix A contains a complete bibliography of published sources cited throughout the *Guide*.

- 0.12 Controversy is endemic to scientific inquiry, which proceeds by adopting an attitude of organised scepticism and perforce challenges orthodox beliefs and assumptions. There naturally remain areas of DNA profiling, in both theory and practice, that are subject to uncertainty and competing interpretations by specialists. Moreover, even if a particular test result, statistic, or probabilistic calculation is undeniably sound, its potential *forensic applications* (including the threshold question whether it should have any forensic applications at all) may be matters of fierce debate between proponents and their critics, possibly adopting divergent starting points and assumptions.

The following exposition endeavours to present ‘just the essential facts’ about DNA evidence and its forensic evaluation, as neutrally as possible. Where we occasionally

found it impossible or inappropriate to steer clear of all controversy, we have tried to indicate the range of potential alternative approaches and to assess 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 promoting more fully informed uses and evaluations of DNA evidence. The overarching goal of all four *Guides* is to improve the quality of probabilistic reasoning and to facilitate the communication and interpretation of probabilities and statistical evidence in the administration of criminal justice.

0.13 The preparation of this *Guide* has benefited enormously 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 stakeholders. Roberto Puch-Solis and Susan Pope thank their FSS colleagues, Samantha Underwood, Jon Wetton, Valerie Tucker, Andrew Hopwood and Ian Evett for many helpful discussions on the subject matter of this *Guide*. We are especially grateful to HHJ John Phillips and to Sheriff John Horsburgh for their support and comments. 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 four named authors, and none of our Working Group colleagues or other advisers and commentators should be assumed to endorse all, or any particular part, of our text.

The vital contribution of the Nuffield Foundation, without whose enthusiasm and generous financial support this project could never have been brought to fruition, is also gratefully acknowledged. 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. Whilst the Nuffield Foundation is our primary funder, the named authors take sole responsibility for the views expressed in this *Guide*, which are not necessarily endorsed by the Foundation. More information is available at [www.nuffieldfoundation.org](http://www.nuffieldfoundation.org).

- 0.14 We welcome further constructive feedback on all four published and planned *Guides*. We are keen to hear about practitioners' experiences of using them and to receive suggestions for amendments, improvements or other material that could usefully be incorporated into revised editions.

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

Alternatively, responses by email may be sent to [c.g.g.aitken@ed.ac.uk](mailto:c.g.g.aitken@ed.ac.uk), with the subject heading 'Practitioner Guide No. [1 and/or 2, as appropriate]'.

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.

Roberto Puch-Solis  
Paul Roberts  
Susan Pope  
Colin Aitken

February 2012

## **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

Mark Fenhalls, Criminal Bar Association

Nina Fisher, Law Society of England and Wales

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

Peter Gill, University of Oslo

HHJ Anna Guggenheim QC

Graham Jackson, Abertay University and the Forensic Science Society

John Logue, Crown Office

Tim Moloney, Criminal Bar Association

Tom Nelson, Scottish Police Services Authority

Niamh Nic Daeid, University of Strathclyde

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

### ***International Advisory Panel***

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. DNA Evidence in Criminal Proceedings

## 1.1 *Potted History*

DNA evidence has come to epitomise scientific proof in legal proceedings. There can hardly be any potential juror in the country who has not heard the term ‘DNA profiling’ or is unaware of the almost miraculous potential of ‘DNA’ (further specification has become superfluous) to solve crimes and lead to the conviction of the guilty. DNA has also played a pivotal role in exonerating the wrongly convicted, including death-row inmates in the USA (Connors et al., 1996). The UK established the first national DNA database (NDNAD) which, following a period of planned and well-resourced expansion (McCartney, 2006), is currently the second largest in the world, after the FBI’s CODIS system in the USA. DNA profiling is not confined to the investigation and proof of serious crimes like homicide, rape and armed robbery, but also features in prosecutions of more routine ‘volume’ crimes like domestic burglaries, car theft and street-level drug-dealing.

Today, DNA is widely described as the ‘gold standard’ of scientific evidence (Lynch, 2003). It is also central to debates over the so-called ‘CSI effect’, whereby lay juries are supposedly influenced in their evaluation of evidence by misconceptions or inflated expectations engendered by popular fictional portrayals of the amazing power of science in the administration of criminal justice (see e.g. Schweitzer and Saks, 2007; cf. Cole and Dioso-Villa, 2009).

1.2 DNA evidence has come a remarkably long way in a comparatively short period of time. Following its invention and first tentative forensic applications in the mid-1980s, DNA profiling was subjected to quite intensive legal scrutiny and underwent various technical refinements, some of them in direct response to problems or concerns identified in criminal trials and appeals.

The technology of DNA profiling, which is outlined in Part 2 of this *Guide*, has advanced considerably since its first experimental applications. The basic techniques for producing a DNA profile are now regarded by experts as tried-and-tested, and are rarely challenged in criminal proceedings or anywhere else. This does not imply,

however, that the incremental process of technical refinement and adjustment has come to a halt. We see continuing advances, and on-going controversy, in relation to profiles generated from very small amounts of DNA and regarding ‘mixed profiles’ containing the DNA of more than one donor, for example. These are some of the topics canvassed in the following pages.

### 1.3 *The Probabilistic Foundations of DNA Evidence*

One of the most distinctive features of DNA profiling, as compared with older and hitherto more established branches of forensic science and forensic medicine, is that DNA evidence is explicitly probabilistic. An expert witness does not – or at any rate, certainly should not – identify a particular individual as the donor of the genetic material from which a DNA profile was produced. As Part 2 explains more fully, this is because the standard DNA profile is produced from only a small sample of the donor’s entire DNA. Thus, even if DNA itself is assumed to be unique to each individual, more than one person could still share the same DNA *profile*, e.g. more than one person could be ‘a match’ to crime scene DNA.

In *Doheny and Adams* the Court of Appeal remarked that, ‘[a]s the art of analysis progresses... the stage may be reached when a match will be so comprehensive that it will be possible to construct a DNA profile that is unique and which proves the guilt of the defendant without any other evidence. So far as we are aware that stage has not yet been reached’.<sup>7</sup> In fact, DNA profiling will *never* be able to produce a verifiably unique match to a particular individual, because the evaluation of DNA evidence is always, in part, a question of probability.

1.4 The overtly probabilistic foundations of DNA profiles have important implications for the production, presentation and evaluation of DNA evidence. In the first instance, the forensic scientist must arrive at her own assessment of the results of DNA profiling and their potential probative value as evidence – bearing in mind that the fact-finder is always the ultimate arbiter of probative value in criminal trials. The forensic scientist must then communicate this information effectively to the instructing prosecutor and prosecuting counsel (or to defence lawyers, as the case may be) in preparation for

---

<sup>7</sup> *R v Doheny and Adams* [1997] 1 Cr App R 369, 373.



trial. If lawyers do not understand the meaning or probative value of DNA profiles there is a danger that the expert's evidence will be misunderstood and then misrepresented in court. Counsel might ask the wrong question or inadvertently invite the expert to rephrase her evidence in a misleading or obscure way, leaving the jury confused and in the dark. This partly explains why experts sometimes fall into errors like the much-discussed '*prosecutor's fallacy*' (which is revisited in Part 7) whilst testifying in the witness-box.

Moreover, in the potentially confrontational context of an adversarial criminal trial, some element of confusion might be part of the opposing advocate's deliberate strategy in cross-examination. An expert witness with a thorough grounding in DNA's probabilistic foundations will be better prepared to resist distortions of her evidence, and to avoid saying the wrong thing in the heat of the moment, than one whose grasp of the probabilistic and statistical dimensions of DNA evidence was tenuous to begin with.

Finally, even if the expert's evidence is accurate and clear, there remains the challenge of successfully communicating the true probative potential of DNA evidence to the trial judge and to jurors. Lay jurors are likely to need some guidance in making sense of evidence expressed in terms of probabilities. This has implications both for the way in which DNA evidence is presented and tested in court, and also for the way in which trial judges sum-up DNA evidence for the benefit of the jury. An effective summing-up presupposes that trial judges themselves properly understand DNA evidence. In addition, as Part 7 explains, it is possible to communicate the probabilistic aspects of DNA evidence in a variety of ways, employing different expressions and formulations. UK courts have been drawn into controversies over the most appropriate method for presenting DNA evidence in the courtroom, leading to important rulings on admissibility, especially by the Court of Appeal in relation to English law.<sup>8</sup>

---

<sup>8</sup> Notably including the three 'Adams family' decisions: *R v Adams* [1996] 2 Cr App R 467, CA; *R v Doheny and Adams* [1997] 1 Cr App R 369, CA; *R v Adams (No.2)* [1998] 1 Cr App R 377, CA; see Roberts and Zuckerman (2010: 159-163).

These are some of the central issues explored in this *Guide*.

- 1.5 Probability calculations were originally widely perceived in legal circles as a special characteristic of DNA evidence, whilst other areas of forensic expertise – notably including fingerprinting – claimed to be able to identify individuals as the unique source of physical evidence (Cole, 2005). In fact, DNA evidence is the true exemplar. The ensuing decades have witnessed a growing realisation that *all scientific evidence is probabilistic* and no current forensic technology supports unique identification of individuals. DNA is different only insofar as it wears its probability on its sleeve, whereas other sciences and technologies have tended to conceal their probabilistic foundations in ostensibly binary concepts such as ‘match’/‘no match’. Forensic scientists have begun to address this transparency deficit in recent years (see, e.g., National Research Council, 2009; The Fingerprint Inquiry: Scotland, 2011).

It is in this sense that Saks and Koehler (2005) proclaimed DNA the model for a ‘paradigm shift in the traditional forensic identification sciences in which untested assumptions and semi-informed guesswork are replaced by a sound scientific foundation and justifiable protocols’. In particular, they urged, ‘[w]hen matches are identified, forensic scientists in all fields would compute and report random-match probabilities similar to those used in DNA typing’. This would (at least in theory) assist the fact-finder to interpret and better evaluate scientific findings expressed in the traditional language of ‘match’, ‘no match’, etc. The ‘new paradigm’ thesis has generated controversy within the wider forensic science community. But if Saks and Koehler are even only half-right, the practical ramifications of the issues explored in this *Guide* are unlikely to be confined in the coming months and years exclusively to the evaluation of DNA evidence in criminal proceedings.

#### 1.6 ***Putting DNA in its Probative Place***

Before launching into discussion of the probabilistic foundations of DNA evidence, it is worth reiterating some basic general propositions about the nature of evidence in criminal trials. These elementary principles frame this *Guide*’s analysis and should be borne in mind as the discussion proceeds.

1.7 First and foremost, it needs to be remembered that evidence cannot be adduced in criminal trials unless it is relevant to a fact in issue. Relevance is the first hurdle to admissibility. It is therefore essential to pay close attention to the fact or facts in issue that DNA evidence purports to prove.

DNA evidence is virtually always adduced as proof of identity, the identity of the offender, of the victim, or of some other individual pertinent to the inquiry. The identity of the culprit is sometimes *the* key issue in the case, as where the accused claims ‘mistaken identity’ or advances an alibi. Un-witnessed homicides, burglaries and rapes by strangers often fit this evidential pattern.

But there are many other commonplace scenarios in which, by contrast, the probative value of DNA evidence as proof of identity would be radically reduced or even completely eliminated. If the accused is admitting sexual intercourse and asserting that the complainant consented, DNA evidence will not provide much assistance to the prosecution in proving a charge of rape. Likewise, if the accused previously had legitimate access to the property, DNA collected from inside the property will not be evidence of burglary; unless the DNA was recovered from a place or in circumstances suggestive of criminal activity, e.g. from blood around a broken window used by the burglar to gain unlawful access. Again, DNA evidence does not rebut a claim of self-defence to a charge of assault, unless there are circumstantial details inconsistent with the accused’s account. And so on. In practice, such evidence would still always be adduced, if only as an agreed statement – not least because the accused might otherwise resile from previous admissions – but its contribution to proving the prosecution’s case at trial will be minimal.

These observations are truistic, but the threshold requirement of relevance to a fact in issue may easily be overlooked, especially in the midst of involved debates about the probabilistic foundations of DNA evidence.

1.8 It is sometimes claimed or implied that DNA evidence is capable of supplying the entire evidential basis of a criminal conviction in England and Wales (though not currently in Scotland, where corroboration is still usually required for criminal

convictions).<sup>9</sup> However, this is never literally true. At the very least, there will also be evidence of what the accused said at or before his trial, or evidence that he said nothing throughout (which is additional information available to the fact-finder, whether or not it qualifies as ‘evidence’ in the technical legal sense). Often, the jury will have access to much else besides.

So when it is said that DNA was the sole basis for a criminal conviction, we should understand this to mean that a DNA profile was central, and quite possibly vital, to the prosecution’s case, without losing sight of any other significant information communicated to the jury. This is simply a contextualised application of the more general point about relevance and probative value highlighted in the previous paragraphs.

- 1.9 Like all physical evidence, DNA profiles present issues of provenance and continuity of physical samples. The probative value of a DNA profile, quite irrespective of its notional weight, hinges crucially on a series of prosaic assumptions, including the following: (i) genetic material from which a DNA profile could be generated remained at the crime scene, without irremediable degradation or contamination; (ii) the physical sample was collected properly at the crime scene (or from the suspect, victim, or whatever); (iii) the sample was successfully transported to the laboratory without interference or contamination; (iv) at the laboratory the sample was analysed using appropriately calibrated and properly functioning machinery, in accordance with appropriate scientific protocols; (v) the results of the tests were accurately observed and recorded; and (vi) at no stage during laboratory testing procedures did the sample become contaminated with other genetic material, wrongly labelled, switched with other samples, etc.

Over time, various practical solutions have been devised to improve each stage in this process (Lynch et al., 2008: chs 4 & 7). For example, physical samples must be placed

---

<sup>9</sup> The Carloway Review of Scots criminal law and practice has recently recommended the abolition of Scotland’s general corroboration requirement, on the basis that ‘[e]vidence should be about its relevance and quality and not, as is currently predominant in Scots criminal law, its quantity’ (17 November 2011): [www.scotland.gov.uk/About/CarlowayReview](http://www.scotland.gov.uk/About/CarlowayReview).

in tamper-evident packaging and carefully labelled. Scientific tests and protocols undergo extensive validation. Forensic laboratories must operate fastidiously controlled contamination-free environments, and are subjected to declared and undeclared ('blind') trials to demonstrate operational reliability. Ideally laboratory error rates in these trials should be publicly available. Extensive systems of training and accreditation have been developed over the last several years and undergo almost continuous refinement.

These important issues are briefly revisited in para.6.7 of this *Guide*, specifically in relation to Low Template DNA. But it should be clear at the outset that the probative value of DNA evidence is radically undercut if we cannot be confident that samples were uncontaminated, tests were accurate, and – in the extreme case – that the sample from which the DNA profile was generated is the same sample that was collected from the crime scene. Kaye and Sensabaugh (2011: 156) suggest that '[s]ample mishandling, mislabelling or contamination, whether in the field or in the laboratory, is more likely to compromise a DNA analysis than is an error in genetic typing'. Trial judges should rule DNA profiling evidence inadmissible if, when called upon to do so, the party proffering the evidence is unable to establish provenance and continuity to the court's satisfaction (Pattenden, 2008).

#### 1.10 *DNA and the Law of Evidence*

Finally, we should note that, beyond its own distinctive evidentiary characteristics and bespoke regulation, DNA evidence is naturally subject to the general law of criminal evidence and procedure. One important implication (without getting into doctrinal specifics which vary across legal jurisdictions) is that proffered DNA evidence could be excluded from a criminal trial on a variety of grounds, in addition to the considerations of relevance and provenance already mentioned.

Prosecution evidence is sometimes excluded on grounds of minimal probative value, or where its probative value – even if substantial – is outweighed by its potentially prejudicial effect on the fairness of the trial. Hearsay and bad character evidence are good illustrations of types of evidence which have traditionally been subject to this kind of admissibility regime in common law jurisdictions.

1.11 Alternatively, evidence may be excluded on the basis of procedural impropriety that has no direct bearing on the probative value of the evidence. For example, if a DNA sample were procured through serious police illegality or outrageous investigative practices, the criminal courts might refuse to receive it simply on the basis that the evidence is tainted and incompatible with judicial integrity and the right to a fair trial.

UK courts have not been particularly receptive to the argument that DNA evidence should be excluded because it derives from an illegally obtained, or illegally retained, suspect profile. They have preferred to overlook illegalities rather than excluding highly probative DNA evidence from criminal trials.<sup>10</sup> However, there is European case-law which could conceivably apply,<sup>11</sup> especially where reliance is placed on physical samples procured by police or security services overseas, who do not necessarily adhere to the same standards of professionalism or restraint as the modern British police.<sup>12</sup>

1.12 There are currently no special admissibility rules applicable to expert evidence in either England and Wales or Scotland, beyond the threshold requirements that the witness must be a competent expert;<sup>13</sup> and his evidence must be relevant and helpful to the jury in resolving a fact in issue,<sup>14</sup> and not excluded on general principles (Roberts, 2009). The Law Commission (2011) has recommended the adoption of a general reliability test for the admissibility of expert evidence, reminiscent of the well-known *Daubert* standard applied in many legal jurisdictions in the USA,<sup>15</sup> but it remains to be seen whether this proposal will make any headway with policymakers.

DNA profiling was integrated into routine criminal investigations and prosecutions across the UK without any statutory framework to authorise or facilitate its

---

<sup>10</sup> *Attorney General's Reference (No 3 of 1999)* [2001] 2 AC 91, HL.

<sup>11</sup> *Gäfgen v Germany* (2011) 52 EHRR 1; *Jalloh v Germany* (2007) 44 EHRR 32.

<sup>12</sup> Cf. *A v Secretary of State for the Home Department (No 2)* [2006] 2 AC 221, [2005] UKHL 71; *Selmouni v France* (25803/94) (2000) 29 EHRR 403.

<sup>13</sup> *R v Atkins* [2010] 1 Cr App R 8, [2009] EWCA Crim 1876; *R v Stockwell* (1993) 97 Cr App R 260, CA; *R v Silverlock* [1894] 2 QB 766, CCR.

<sup>14</sup> *R v Turner* [1975] 1 QB 834, CA.

<sup>15</sup> *Daubert v Merrell Dow* 125 L Ed 2d 469; 113 S Ct 2786 (1993).

admissibility.<sup>16</sup> Nonetheless, the validity of scientific techniques may be challenged from time to time, and novel forms of evidence or innovative applications of established techniques are especially likely to attract adversarial objections and closer judicial scrutiny. Low Template DNA, discussed in Part 6 of this *Guide*, has recently fallen into this category.

In the absence of a dedicated admissibility rule governing novel scientific evidence, judicial determinations of admissibility turn on threshold judgments of relevance. If there are very serious doubts about the validity or reliability of a particular scientific technique, which the party seeking to adduce the evidence cannot dispel to the court's satisfaction, a trial judge might well conclude that the evidence is incapable of assisting the jury to determine the facts in issue on a rational basis: in other words, that the evidence is irrelevant and therefore inadmissible.<sup>17</sup> This is a threshold and fairly undemanding standard of admissibility. Any objection to scientific evidence falling short of outright invalidity or complete unreliability will generally be regarded as a matter of weight only, to be argued in the trial and assessed by the jury.<sup>18</sup>

---

<sup>16</sup> A succession of important changes have, however, been made to the law of criminal procedure to enable the police to take physical samples from suspects, without their consent, from which DNA profiles can then be generated. Most recently, see Crime and Security Act 2010, amending PACE 1984, ss.61-64.

<sup>17</sup> *R v Broughton* [2010] EWCA Crim 549.

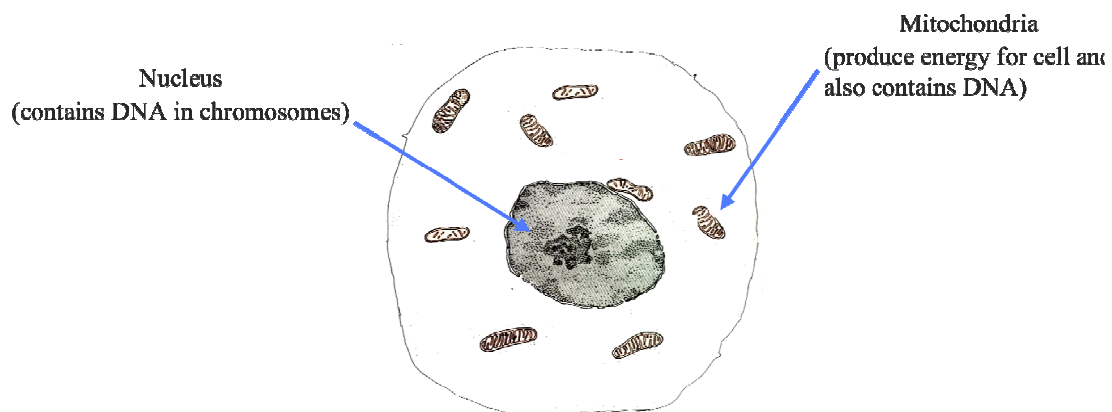
<sup>18</sup> *R v Atkins* [2010] 1 Cr App R 8, [2009] EWCA Crim 1876; *R v Kempster (No.2)* [2008] 2 Cr App R 19, [2008] EWCA Crim 975; *R v Robb* (1991) 93 Cr App R 161, CA.

## 2. The DNA Profile

2.1 This Part briefly explains the biology of DNA and the technology of DNA profiling, as a basis for investigating the probative value of DNA evidence in the remainder of the *Guide*.

### 2.2 **Basic Biology**

**Cells** are the microscopic building blocks, and smallest working unit, of any living organism. They typically consist of a liquid called **cytoplasm**, which contains the instructions for reproducing the chemical ‘machinery’ running the cell, and a **nucleus**, all of which are contained within an external cell **membrane** (see Fig. 2.1). All human cells, with the exception of mature red blood cells, have a nucleus which contains a set of molecules called **chromosomes**.



**Figure 2.1: Diagrammatic representation of a human cell**  
(public domain images from Wikipedia)

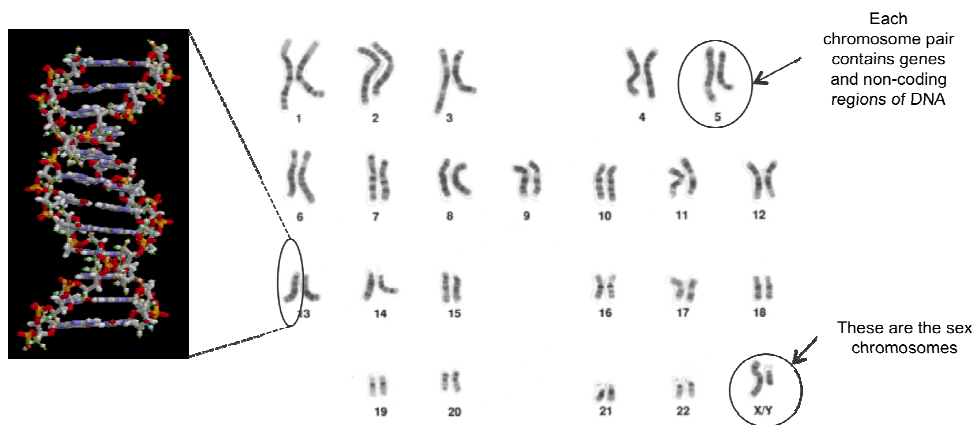
2.3 Typically, humans have 23 pairs of chromosomes, including one pair comprising the two sex chromosomes (X/Y in Fig. 2.2, below). Females have two X chromosomes, whilst males have one X and one Y chromosome. Chromosomes determine a person’s physical characteristics and regulate chemical processes in the human body. One chromosome in each pair is inherited from the father, and the other from the mother. The maternally inherited chromosome is formed of sections of the mother’s two



chromosomes. Likewise, the paternally inherited chromosome is formed of sections of the father's two chromosomes.

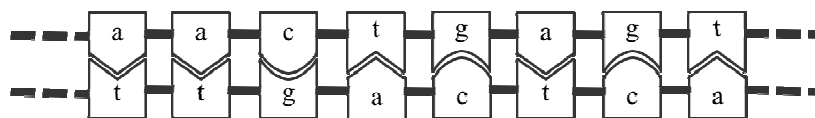
Chromosomes have *coding* and *non-coding* regions. Certain portions of chromosomes are also known as *genes*. The coding regions are the parts of genes which determine a person's physical characteristics.

2.4 Each chromosome is a deoxyribonucleic acid (DNA) molecule. The DNA molecule consists of two strands coiled around each other, forming the characteristic *double helix* (Fig. 2.2).



**Figure 2.2: A diagrammatic representation of human chromosomes**  
(public domain images from Wikipedia)

DNA is formed from four chemical 'bases' called adenine, cytosine, guanine, and thymine. These bind together in pairs within the double helix according to a strict regular pattern. Base *a* binds only with base *t*, and base *g* binds only with base *c*, as shown in Figure 2.3.



**Figure 2.3: Diagrammatic representation of the base pairs in DNA**

2.5 For forensic applications in particular, the length of a section of DNA is measured in terms of its number of *base pairs*. For example, the length of the DNA fragment represented in Figure 2.3 is eight base pairs.

2.6 Cellular cytoplasm also contains *mitochondria* (see Fig 2.1), which are the cell's energy source. Each mitochondrion contains a small circular DNA double helix, referred to as *mitochondrial DNA* and often abbreviated to *mtDNA*. Mitochondrial DNA must be distinguished from nuclear DNA (sometimes also referred to as 'chromosomal' DNA). Whereas nuclear DNA is inherited from both parents, mtDNA is inherited only from the mother. Forensic applications of mtDNA are discussed further in Appendix D.

### 2.7 *Profiling Nuclear DNA*

The DNA of each individual comprises millions of base pairs. It is neither feasible nor necessary for forensic profiling to attempt to reconstruct a person's entire DNA. Instead, forensic profiles sample a small number of regions of DNA, known as *loci* (singular, locus). Non-coding regions exhibit far greater variation than coding regions, making them particularly suitable for forensic DNA profiling. Greater variation increases the technique's power to discriminate between individuals. In addition, non-coding regions have no apparent observable effects on human characteristics, thus reducing concerns about privacy and medical confidentiality (e.g. in relation to genetic illness).

2.8 Specifically, forensic DNA profiles express values for *short tandem repeats* (STRs), which are short sequences of base pairs repeated multiple times. The number of times that the sequences are repeated varies greatly between individuals. The length of each repeated sequence can be measured and expressed as the number of repeats in the sequence. This is called an *allele*.

Fig. 2.4 illustrates a locus where the sequence 'g-a-t-a' is repeated 4 times. So this person would be said to have 'allele 4' at this particular locus.

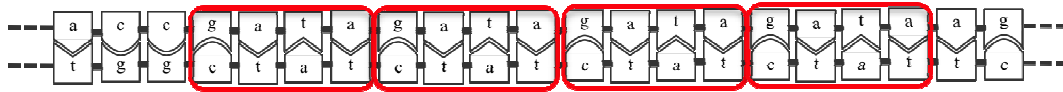


Figure 2.4: An illustration of allele 4 at a locus

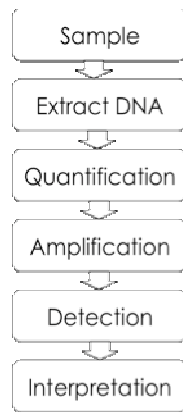
2.9 A **genotype at a locus** consists of two STR values, one for the allele inherited from the father and one for the allele inherited from the mother. If these alleles have two different values, the person is said to be **heterozygous** at that locus. If the values for the two alleles happen to be the same, the person is **homozygous** at that locus.

During the natural process of cell division, DNA molecules occasionally fail to reproduce themselves accurately, leading to **mutations**. This might involve adding, changing or deleting one or more base pairs at particular loci. Genetic mutations can have serious implications for an individual's health and wellbeing, but they are irrelevant for most forensic purposes, given the restriction of forensic profiling to non-coding regions with no known functional effects. If a DNA profile produced from a crime stain contains a mutation, the true donor's profile will also contain the same mutation, and there will be no particular difficulty in linking this unusual crime stain profile to its donor. However, mutations may need to be considered if genetic inheritance is in issue, e.g. where paternity is contested.

2.10 **Obtaining an STR DNA profile**

Once a biological sample has been obtained, there are five key stages in the forensic DNA profiling process

This series of sequential steps, most of which can be automated, is represented diagrammatically by Figure 2.5:



**Figure 2.5 The process for obtaining DNA profiles**

**2.11 (i) Extraction from sample**

Profiling begins with a sample of biological material, such as blood, semen, saliva, hair or skin cells, which may have been recovered from a crime scene, from a victim, or from a suspect, etc, and submitted to the laboratory for testing. The nuclear DNA contained in these samples must be purified by extracting it from the cells prior to any further analysis.

**2.12 (ii) Quantification**

Profiling technologies are sensitive to the quantity of DNA tested. If there is either too little or too much DNA present in the sample the test is likely to fail. The amount of extracted DNA must therefore be measured to ensure that it falls within an appropriate range. The typical amount of DNA used in profiling is between 150 picograms<sup>19</sup> and one nanogram.<sup>20</sup>

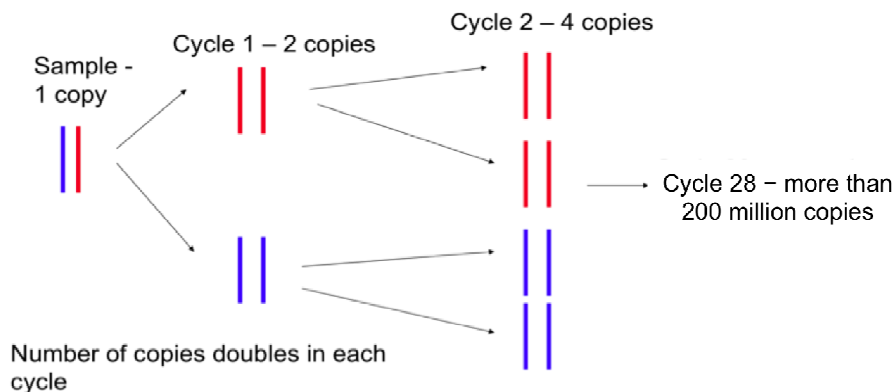
**2.13 (iii) Amplification**

The amount of DNA extracted from a forensic sample is too small to be detected by standard profiling equipment and techniques, and therefore needs to be increased through a process known as *amplification* - often referred to as the *Polymerase Chain Reaction* or PCR. In this step the DNA forming STRs at designated loci is duplicated many times over, as indicated by Figure 2.6:

---

<sup>19</sup> A picogram is  $10^{-12}$  grams.

<sup>20</sup> A nanogram is  $10^{-9}$  grams.



**Figure 2.6 DNA duplication in the amplification step**

The amount of DNA available for testing doubles in each copying cycle. The number of cycles employed in the amplification process is determined by the particular *DNA profiling system* being used. The current UK standard system stipulates 28 or 34 cycles (NPIA, 2010).

2.14 During amplification, the DNA fragments are also chemically labelled by adding a light-reactive dye that can be detected later in the process.

2.15 **(iv) Detection**

At the detection stage, each sample of (now labelled) DNA is transmitted through a separate capillary until it reaches a laser. The laser causes the chemical labels on the DNA fragments to fluoresce. These light emissions are detected by a scanner and recorded by computer.

2.16 **(v) Interpretation**

Finally, dedicated software is used to interpret the computer-generated data. The intensity and position of each light emission, displayed as a peak on an *electropherogram* (EPG), is compared against standardized measures of known size and amount. Peak heights are measured in *relative fluorescence units* (rfu).

2.17 **Protocols, multiplexes and systems**

DNA profiling may employ a range of different equipment and settings, which in turn have their own particular interpretative guidelines. For example, increasing the number of amplification cycles from 28 to 34 cycles affects the behaviour of the peak heights in the profile. The list of equipment and settings used in a specified process for producing a profile is known as a **protocol**. Even in a largely automated technological process, human judgement – the skill and experience of the forensic scientist, applying validated laboratory procedures – enters into the practice of successful profiling and influences the interpretation of its results.

2.18 Forensic DNA profiles consist of collections of STR loci that are analysed together at the same time. Such collections of loci are called **multiplexes**. There are various multiplexes available from commercial suppliers, each of which tests a particular set of loci. Figure 2.7 lists some of the more widely used multiplexes, and indicates their distinctive sets of loci.

**Figure 2.7 Examples of multiplexes**

Locus	SGMPlus*	ESS**	ESI***	Identifiler	CofilerPlus
D3	•	•	•	•	•
vWA	•	•	•	•	•
D16	•		•	•	
D2q	•		•	•	
D8	•	•	•	•	•
D21	•	•	•	•	•
D18	•	•	•	•	•
D19	•		•	•	•
TH01	•	•	•	•	•
FGA	•	•	•	•	•
D1		•	•		
D2p		•	•		
D10		•	•		
D12		•	•		
D22		•	•		
D7				•	•
CSF1PO				•	
D13				•	•
TP0X				•	
D5				•	
Amelo	•	•	•	•	•

\*Second Generation Multiplex Plus (Applied Biosystems); \*\* European Standard Set;

\*\*\* European Standard Identifier (Promega)

Loci are generally known by their abbreviations, as shown in Figure 2.7, rather than writing out in full their long and complicated chemical names. A *DNA profiling system* comprises a multiplex and a protocol.

2.19 In England and Wales, the Custodian of the National DNA Database (NDNAD) has specified requirements for the multiplexes that forensic service providers (FSPs) must use in order to upload profiles to the NDNAD. The currently validated standard is the SGMPlus multiplex, with 28 or 34 amplification cycles (NPIA, 2010).

There are on-going efforts to standardise DNA profiling systems, not least to facilitate international data-sharing, cross-border policing and mutual judicial assistance. EU member states, including the UK, committed themselves to adopting multiplexes covering the European Standard Set of loci (ESS in Figure 2.7) by November 2011 (EU Council, 2009).

2.20 For purposes of illustration, this *Guide* will take the SGMPlus profiling system as standard, but our analysis applies *mutatis mutandis* to any other forensic multiplex.

2.21 Figure 2.8 provides an example of an EPG generated from the SGMPlus profiling system. This is the tangible final product of the procedure described in paras.2.10 – 2.16, above.

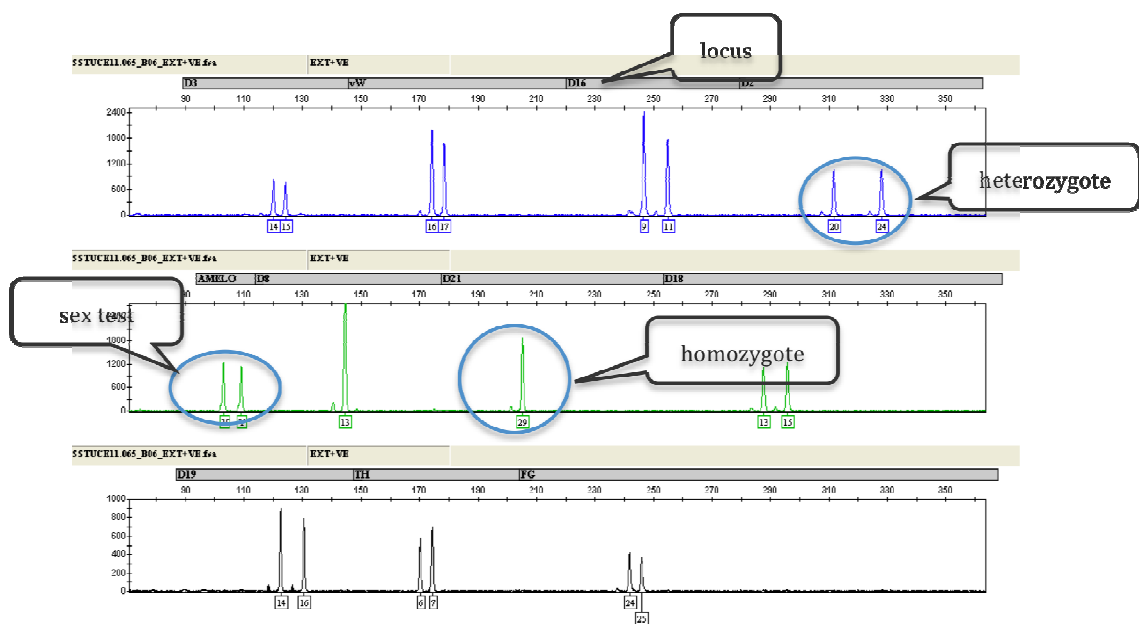


Figure 2.8 EPG of an SGMPlus DNA profile (10 loci + sex test) from one person

At each multiplex locus there are either one or two recorded peaks, corresponding to the donor's alleles at that locus. At locus D2, for example, there are two peaks (with STR values 20 and 24, respectively), meaning that the donor is heterozygous at locus D2. But there is only one peak at locus D21, with STR value 29, so the donor is homozygous at this locus.

2.22 All current multiplexes include a sex test. The donor of this sample profile must be male, since both X and Y chromosomes have been detected. There is also a type of DNA test that considers only the Y chromosome, which is discussed further in Appendix C.

2.23 As previously noted, analysing DNA samples using different multiplexes and protocols affects the height of the peaks recorded by the EPG. However, the basic principles summarised in these paragraphs apply equally to all multiplexes and protocols.

2.24 The data contained in forensic DNA profiles can be conveniently recorded, computerised, stored and searched, as pairs of numbers (corresponding to STR values for alleles) for each locus in the multiplex. Thus, the EPG depicted in Figure 2.8 can be translated directly into Figure 2.9:

D3	vWA	D16	D2	AMELO	D8	D21	D18	D19	TH01	FG
14,15	16,17	9,11	20,24	X,Y	13,13	29,29	12,15	14,16	6,7	24,25

**Figure 2.9 SGMPlus DNA Profile from one person**

2.25 The reader should by now have a reasonably clear picture (in words and pictures) of what a forensic DNA profile actually is. We now need to ask: what is the (potential) probative value of a DNA profile in the context of criminal adjudication? What weight should the fact-finder give to this evidence, in light of the disputed matters in issue in the proceedings and the totality of the evidence adduced in the trial?



### 3. DNA Profiles as Evidence in Criminal Proceedings

3.1 Part 2 summarised the basic science and technology involved in the production of a forensic DNA profile. This Part draws attention to some of the additional considerations and complexities involved in translating the theory of DNA profiling into the realities of criminal justice practice.

#### 3.2 *Source, classification and quality of samples*

As *Practitioner Guide No 1* explained, there are two types of samples that are routinely considered in the production and evaluation of forensic science evidence (Aitken, Roberts and Jackson 2010: 24). Samples of unknown origin (e.g. blood recovered from a crime scene, or semen recovered from a rape victim) are known as *questioned* (or recovered or crime) samples, whilst samples of known origin are called *reference* samples.

By extension, a DNA profile obtained from a sample is either a *questioned profile* (when obtained from a questioned sample) or a *reference profile* (when obtained from a reference sample).

3.3 Reference samples are obtained deliberately and under controlled conditions, often by taking a buccal swab from a suspect at a police station. A reference sample can be expected to contain substantial amounts of good quality DNA to facilitate profiling.

The EPG of a reference profile obtained under these ideal conditions consists of large peaks, such as those seen in Figure 2.8, which enable the alleles present at each locus to be detected successfully. A person's reference profile is also known, in the context of forensic DNA profiling, as that person's *genotype*.

3.4 By contrast, the quality and quantity of DNA in a questioned sample can vary enormously. Various factors are at work.

3.5 First, environmental conditions can affect the quality of a profile. For example, samples exposed to rain or submerged in water (e.g. blood on a weapon recovered from a canal) are liable to deteriorate. DNA extracted from such samples would

probably degrade. Generally speaking, hot and wet conditions cause more degradation than dry and cold conditions.

The presence of chemicals in the sample can also affect the quality of a profile. For example, the dyes in blue jeans can inhibit efficient amplification of DNA.

3.6 A second set of factors affecting the amount of DNA that can be extracted from a sample concern the nature and type of the stain (or other biological material), and the length of contact between that material and the surface from which it was recovered.

For example, more DNA would normally be extracted from a large fresh bloodstain than from skin cells recovered from a tool that was handled only briefly. The nature and duration of contact with an item also has an effect. So, the owner of a handbag would deposit more of her DNA on the handle than a thief who touches it only once. And so on.

3.7 Thirdly, questioned samples may contain mixtures of DNA from multiple donors. Vaginal swabs taken from rape complainants will obviously contain the complainant's DNA as well as the perpetrator's (whomever he may be). Weapons or clothing associated with particular crimes may have been stained by blood from more than one victim or perpetrator, etc.

### 3.8 *Types of questioned profile*

A questioned profile consists of the alleles and peak heights in the EPG. Several types of questioned profile may be distinguished.

3.9 A *full profile* contains all the alleles from a single donor (and thus resembles the genotype produced from a reference sample). In a *partial profile*, at least one allele is missing. A *mixed profile*, by contrast, contains the DNA of more than one donor, and consequently registers more than two alleles at multiple loci.

Figure 3.1 illustrates these various possibilities:

**Figure 3.1 Comparison of reference and questioned profiles**

Locus	Type of Profile				
	1. Reference	2. Full	3. Partial	4. Mixed	5. Different donor
D3	14,15	14,15	14,15	14,15, 16	16,17
vWA	16,17	16,17	16,17	14,16, 17	16,19
D16	9,11	9,11	9	9,10,11,12	10,12
D2	20,24	20,24		20,24	20,20
Amelo	X,Y	X,Y	X,Y	X,Y	X,Y
D8	13,13	13,13	13,13	13,14, 15	13,14
D21	29,29	29,29	29	29,30	29,29
D18	12,15	12,15		12,15	12,15
D19	14,16	14,16	14,16	14,16,20,22	15,15
TH01	6,7	6,7	6	6,7,9.3	6,7
FGA	24,25	24,25		24,25	24,25

The reference and full profiles depicted here contain two, and only two, alleles at each locus. Since the STR values for each allele at all eleven loci are the same, the donor of the reference profile could also be the source of the full profile. Put another way, these results do not exclude the possibility that the donor of the reference profile is also the source of the full profile.

The mixed profile in column 4 of Figure 3.1 comes from at least two people. Notice that there are four alleles at D19 but only two alleles at D18. Profiles from different donors may have alleles in common (here, the two donors share 12,15 at locus D18), but there will virtually always be differences. For example, the alleles at locus TH01 are the same in profiles 1 and 5, but they clearly come from different donors, as can be seen by cross-checking the values at other loci, e.g. D3 where profile 1 registers 14,15, but profile 5 is 16,17.

### 3.10 *Accounting for peak height variation and natural artefacts*

In real world testing, the results of DNA profiling will always contain minor variations in peak heights for profiles produced from the same sample. These are the result of the profiling process. For example, laboratory procedures for extracting DNA from biological material could result in slightly different amounts of DNA being harvested from the same plastic tube. Or the amplification step of the profiling process may be more efficient in one test than in another; and so on. In addition, as

with any laboratory process involving biological samples, some artefactual peaks may be produced as a by-product of the profiling process.

Data have been collected documenting the nature and frequency of variations in profiling results utilising different protocols. Drawing on this research, it is possible to try to take account of natural variations and artefacts in the profiling process. Some of these artefacts are addressed at the interpretation stage (para.2.16, above) while *heterozygote balance* and the *stutter ratio* figure in assessments of probative value.

### 3.11 (i) *Heterozygote balance*

The EPG of a complete profile will indicate two alleles present at every locus where the donor is heterozygous. The peaks forming a heterozygous pair are expected to be roughly, though not exactly, the same height. In general, peak heights become less balanced at the lower end of the range of rfu values.

The degree of balance, or imbalance, between two peaks heights (or ‘areas’) on an EPG can be used to assess whether they do form a genuine heterozygous pair. This is particularly important in relation to possibly mixed profiles, where it is necessary to distinguish between multiple potential donors.

We should emphasise that this *Guide* is intended only to give a broad overview of the concepts and approaches employed in assessing the results of DNA profiling. The Custodian of the National DNA Database, the Forensic Science Regulator and professional organisations such as the International Society of Forensic Genetics ([www.isfg.org](http://www.isfg.org)) produce guidelines and recommendations for specific values and their usage.

### 3.12 Peak height balance can be measured in different ways. The simplest method assesses *heterozygote balance* (Whitaker et al., 2001), calculated according to the formula:

$$\textit{heterozygote balance} = \frac{\textit{height of the shorter allele}}{\textit{height of the taller allele}}$$

Peaks of exactly the same height have a heterozygote balance of one, i.e. perfect balance. The heterozygote balance for a profile produced from an optimal amount of DNA using SGMPlus is expected to be greater than 0.6 for some protocols with 28 amplification cycles, i.e. the height of the shorter allele is at least 60% of the height of the taller allele.

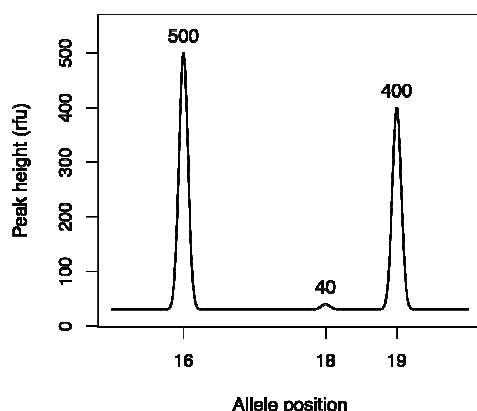
Values can fall below the expected range of heterozygote balance for a variety of reasons. When this occurs, the forensic scientist interprets the result in the light of her assessment of the overall profile, and gauges its probative significance accordingly, drawing on her personal expertise and any relevant data relating to the DNA system being operated.

Heterozygote balance is also one of the interpretive tools for assessing the probative value of mixed profiles, a procedure addressed in Part 5.

### 3.13 (ii) *Stutters*

On some occasions a peak in an EPG occurs with a smaller peak one STR unit to the left. The smaller peak is called the *stutter* of the *parent peak*.

This is shown in Figure 3.2, where the small peak at allele 18 is the stutter of its parent at allele 19.



**Figure 3.2 Part of EPG from Heterozygote Donor**

A stutter is not always detected for every peak in the EPG. Thus, in Figure 3.2 there is no stutter associated with the peak at allele 16.

3.14 Stutters are a natural artefact of the profiling process. It is important to appreciate and account for stutter behaviour when interpreting EPGs, especially in relation to mixed profiles.

The *stutter ratio* (Buckleton et al., 2005) is calculated to assess whether a peak registered in an EPG is likely to be an artefactual stutter rather than a genuine allele:

$$\textit{stutter ratio} = \frac{\textit{height of the stutter peak}}{\textit{height of the parent peak}}$$

For example, the stutter ratio in an SGMPlus profile obtained from an optimal amount of DNA is expected to be smaller than 0.15, or 15%, for some protocols. In Figure 3.2, the measurement of 40 rfu at position 18 is 10% of the parent peak (40/400), suggesting that it could be treated as a stutter. These interpretational questions, requiring forensic scientists to apply their judgement and expertise, are analogous to assessing the implications of heterozygote balance. Alternative approaches currently being developed assess continuous values, without stipulating arbitrarily fixed thresholds for drawing interpretative conclusions, thus making the most of DNA profiling results (Puch-Solis et al., 2012).

### 3.15 *Propositions in the evaluation of DNA evidence*

The preceding paragraphs have described the production and interpretation of EPGs in routine DNA profiling. Where there is a full profile from a single donor it should be quite straightforward to determine whether the genotype of the donor of the questioned sample is the same as the genotype of the donor of a reference sample. In such cases, it makes sense to ask whether the reference sample profile ‘matches’ the profile derived from the crime stain – whether, that is to say, all the alleles in the reference profile are the same as the alleles in the crime stain profile.

However, in cases of partial or mixed profiles or where there are only very small amounts of DNA available for testing, an entirely different approach is required, one which dispenses entirely with the somewhat problematic notion of ‘matching’ profiles. As the next two Parts will explain in detail, this alternative approach assesses

the meaning and probative value of a DNA profile by considering the probability of the evidence under appropriately specified pairs of mutually exclusive propositions. The starting point for further analysis is to consider the level or levels of propositions to which DNA evidence may be addressed in the instant case.

3.16 The idea that evidence may be addressed to different levels of proposition was introduced in *Practitioner Guide No 1* (Aitken, Roberts and Jackson, 2010: §§3.4 – 3.8) and will be explained more systematically by *Practitioner Guide No 4*. Briefly, for present purposes, we can distinguish four basic levels of proposition:

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

This taxonomy combines a mixture of ordinary linguistic usage and more specialist, technical terminology.

Offence level propositions are addressed to whether or not the suspect committed the offence. For example, we might ask whether a confession is probative evidence that the suspect committed murder.

Activity level propositions address whether the suspect (or some other person of interest) performed a relevant action. The action in question could be part of the *actus reus* (conduct elements) of an offence. Thus, we might ask whether the presence of glass on the suspect's clothes is probative evidence that the suspect broke the window. Even if the glass recovered from the suspect's clothes 'matches' (i.e. is indistinguishable from) the glass in the window, it does not necessarily follow that the suspect performed the relevant action – here, breaking the window. He might have walked past the window, contaminating his clothing with glass fragments, shortly after the window was broken by somebody else, for example.

Source level propositions are addressed to the source of particular physical evidence. In relation to biological evidence, source level propositions address whether the

suspect (or other relevant person) is the source of an identifiable body fluid. In this context, ‘body fluid’ essentially means blood, semen, saliva or hair, the latter also being biological material from which DNA may be extracted. Thus, we might ask whether the blood found on the hilt of the knife might have been donated by the suspect, the victim, or some other person.

Finally, sub-source level propositions consider just the DNA in isolation, without attributing it to a particular body fluid. DNA extracted from a crime stain does not necessarily share a common donor with the fluid containing it – as is obvious in cases of mixed profiles, implying two or more donors, derived from a single bloodstain, semen sample, or other bodily fluid.

3.17 The conceptual distinction between (iii) source level propositions and (iv) sub-source level propositions is especially significant in relation to DNA profiling evidence. The crux of the matter can be seen in the difference between saying:

(a) the suspect (or another person of interest) left the body fluid at the scene of the crime (or other location of interest); or

(b) the suspect (or another person of interest) donated the unattributable cellular material from which DNA was extracted.

To ask whether a particular person is the donor of an identifiable body fluid is a source level inquiry. To ask whether a particular person is the donor of DNA extracted from unattributable cellular material is a sub-source level inquiry.

3.18 An important initial consideration when trying to determine the probative value of DNA evidence will therefore always be whether the evidence is probative of source level propositions or only of sub-source level propositions. That is to say, does the evidence purport to identify the donor of a body fluid (source level proposition)? Or, in light of known circumstantial factors affecting the crime or its investigation, are we confined to saying that the evidence purports to identify only the donor of profiled DNA (sub-source level proposition)?



Why is it so important to respect these conceptual distinctions, and to avoid confusing them in thought or expression, during the course of criminal proceedings? One need only consider the difference between an expert witness telling a jury that the victim's blood was found on the accused's clothing, or testifying instead that the victim's DNA was recovered from the accused's clothing. There will often be many perfectly innocent explanations for the presence of DNA which could not plausibly be extended to the presence of blood. A similar distinction between DNA and semen might be equally telling in the context of sexual offence prosecutions, and so on.

3.19 For reasons that the next Part will fully elucidate, propositions in the evaluation of forensic science evidence should always come in pairs: the probability of the evidence in light of the prosecution's proposition must always be compared with the probability of the evidence in light of the defence's proposition (ideally taking account of the arguments to be advanced at trial). Viewed in isolation, the probability of the evidence assuming the prosecution's proposition is uninformative and potentially misleading. Probative value cannot logically be assessed by considering only one half of the equation.

3.20 We have now reached the point in the exposition where the probabilistic nature of DNA profiling evidence can be explored in detail. Part 4 begins with the simplest case, involving a DNA profile with a single donor. Part 5 explores the further issues that must be tackled in relation to mixed profiles with multiple donors, whilst Part 6 examines the technique known as low template DNA profiling that can be applied to samples containing very small amounts of DNA. Part 7 completes our analysis by considering how the evidential fruits of DNA profiling should be communicated to juries in the courtroom, in accordance with legal requirements and the institutional and practical constraints imposed by criminal trial proceedings.

## 4. Assessing the Probative Value of Single Donor Profiles

### 4.1 *The Logic of Probative Value*

The rudiments of a logical approach to assessing probative value were set out in *Practitioner Guide No 1* (Aitken, Roberts and Jackson, 2010: §2.14). Two principles are axiomatic.

First, all empirical propositions are probabilistic. There is no such thing as absolute, complete, unimpeachable and non-revisable certainty in the empirical world. Human decision-making, in other words, occurs under conditions of unavoidable uncertainty. This is clearly reflected in orthodox conceptions of the criminal standard of proof as ‘beyond *reasonable* doubt’,<sup>21</sup> not beyond *all* doubt, or every conceivable doubt, etc.

Secondly, judgements of relevance and probative value are relative. As classical statements of the concept of ‘relevance’ in English law encapsulate, evidence is relevant when it affects the probability that a fact in issue is true or false (Roberts and Zuckerman, 2010: §3.2). Incriminating evidence increases the probability that the accused is guilty. Conversely, evidence which reduces the probability that the accused is guilty – or, in other words, increases the probability that the accused is innocent – is exculpatory. Vitally, evidence that neither increases nor decreases the probability that the accused is guilty is irrelevant and inadmissible, on elementary evidentiary principles.

### 4.2 *The Probative Logic of DNA Evidence*

It is widely believed that DNA is unique to each individual (with the exception that monozygotic biological twins share the same DNA). If true, this is a biological fact with enormous forensic potential. The holy grail of criminal investigation is a form of scientific proof that uniquely identifies an individual as the perpetrator.

---

<sup>21</sup> Juries in England and Wales are now directed that they should be “sure” of the accused’s guilt before convicting. But there is no indication that this was intended to alter the standard of proof, as opposed to being a more effective way of communicating the meaning of the traditional standard to juries.

We have already seen, however, that DNA profiling evidence does not uniquely identify individuals, let alone perpetrators. DNA evidence is generally addressed to source or sub-source propositions, not to offence level propositions. Moreover, even if DNA is unique (biological twins excepted), DNA profiles, produced from only small portions of a person's entire DNA, may not be.

The upshot is that DNA evidence needs to be understood, interpreted and evaluated in expressly probabilistic terms.

4.3 *Practitioner Guide No 1* introduced a logical approach to evaluating evidence in terms of competing propositions (Aitken, Roberts and Jackson, 2010: §2.14). We will now adapt this general framework to the specific requirements of DNA profiling evidence, bearing in mind that, for any evidence to be relevant and admissible in legal proceedings, it must affect the probability that a fact in issue is true or false. DNA evidence is normally adduced as proof of identity, of the offender, the victim, or some other person of interest. Identity is the matter in issue to which DNA evidence must be relevant and probative.

Further analysis proceeds by generating a pair of mutually exclusive propositions (which can also be thought of as competing hypotheses) linking the evidence to contested facts. The following pair of propositions is representative of the way in which the issue might be framed at the sub-source level:

***Prosecution Proposition*** (PP): the DNA recovered from the crime scene came from the accused.

***Defence Proposition*** (DP): the DNA recovered from the crime scene came from an unknown person, not blood related to the accused.

In essence, the prosecution is alleging that the accused is the donor of the DNA, and the defence is denying it. (As we explain in this Part, the defence's denial routinely incorporates the claim that the donor is also not one of the accused's close blood relatives.) Our task is to assess whether DNA profiling evidence affects the respective

probabilities of PP and DP, and if so, by how much. DNA evidence is relevant, probative and admissible just in so far as it can discriminate between PP and DP, by making one proposition more likely to be true and the other proposition less likely to be true.

- 4.4 It bears emphasis that, although probabilistic calculations are integral to the production of DNA profiling evidence, the process of translating biological data into proof of facts in law begins, not with probability or statistics of any kind, but with non-mathematical logical analysis. Relevance and probative value are relative, not absolute, concepts. They invite the further question: relative to (or probative of) *what?*

DNA profiles are assessed by reference to a pair of competing propositions formulated by the forensic scientist (or anybody else undertaking a similar evidentiary assessment), following established protocols and utilising her case-work experience and knowledge of the instant case. Choice of propositions is necessarily determined by the nature of the charges, the prosecution's factual allegations, and whatever is known of the accused's replies and any other significant evidence in the case. Propositions appropriate to one type of charge might fail to address the matters at issue in a different type of case. Even in relation to the same proceedings, propositions might need to be updated and probative value recalculated if salient facts or assumptions change; if, for example, the accused advances a legitimate reason for his presence at the scene, or claims that the crime was committed by his brother.

But without first formulating an appropriate pair of propositions, or 'hypotheses', for PP and DP, analysis of the probative value of DNA evidence cannot even begin.

- 4.5 These introductory remarks clarify the questions that must be answered in our inquiry. First, there is the question of relevance (also the first hurdle to legal admissibility): does DNA profiling evidence affect the respective probabilities of PP and DP?

Secondly, there is the question of probative value: how can we measure the size of the impact of DNA profiling evidence on the probabilities of PP and DP?

4.6 The answer to the first question is that DNA evidence is relevant, and indeed highly probative in many factual scenarios, because it is most unlikely that two people (other than monozygotic twins) would share the same full profile. Thus, if a questioned profile matches a person of interest, this is highly probative evidence that the person of interest is the donor of the DNA from which the profile was generated.

But to say that it is unlikely, or even very unlikely, that two people would share the same DNA profile is clearly not the same as saying that a shared profile is impossible. Even if DNA evidence increases the probability of the prosecution's proposition (PP) enormously, the probability of PP will never equal 1 (100% certainty) and the probability of DP will never drop to zero.

4.7 The answer to the second question lies in the relative frequency of a particular genotype in some relevant population. A characteristic that is commonplace, with a large relative frequency in a given population (say, people in London who speak English), is not very helpful in discriminating between individuals. If all we knew about the perpetrator of a London murder was that the murderer spoke English, we would not know very much at all. Only a very small number of non-English speaking London residents would be ruled out of the pool of potential suspects. Characteristics with small relative frequencies in a given population (e.g. people in London who speak Inuit or Micronesian languages) are much more helpfully discriminating, from a forensic point of view. The vast majority of London residents would be excluded from the suspect pool if a tape-recording of the murderer's voice revealed him making Inuit death threats to the victim.

Human DNA is shared by 100% of humans and, consequently, finding that a questioned sample contains human DNA does not discriminate between a particular suspect and every other human being on the planet. However, we know that the STR values for human DNA vary greatly between individuals, and it is this variation that DNA multiplexes profile, as explained in Part 2. The probative value of a DNA match is then amenable to probabilistic calculation, drawing on statistical data (including estimates of the relative frequencies of particular genotypes within ethnic subpopulations).

#### 4.8 *DNA Likelihood Ratios*

*Practitioner Guide No 1* explained how the probative value of evidence could be assessed as the **likelihood ratio** of the probabilities of the evidence in the light of pairs of mutually exclusive propositions, such as PP and DP (Aitken, Roberts and Jackson, 2010: §2.17). We will now apply this general approach specifically to DNA profiling evidence.

4.9 In relation to DNA evidence (e.g. a profile matching the accused), the likelihood ratio is the ratio of the following two probabilities:

$p(E | PP)$ : the probability of the DNA evidence if the prosecution's proposition is true; and

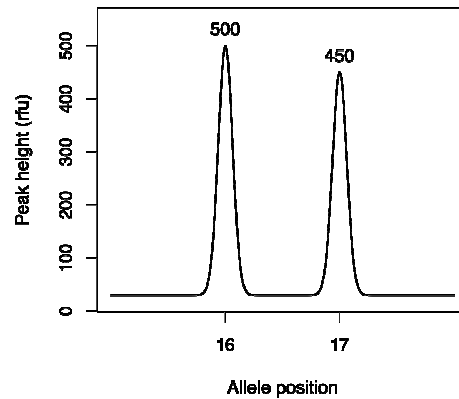
$p(E | DP)$ : the probability of the DNA evidence if the defence's proposition is true.

The first probability gives the numerator of the likelihood ratio, and the second probability supplies the denominator. Thus, if  $p(E | PP) = 0.1$ , and  $p(E | DP) = 0.01$ , the likelihood ratio would be  $0.1/0.01 = 10$ .

A likelihood ratio greater than one supports the prosecution's proposition. A likelihood ratio smaller than 1 would support the defence's proposition. If the likelihood ratio were exactly 1, the evidence would be equally likely under either proposition: that is to say, it would not alter the probability of guilt or innocence, and would therefore be irrelevant and inadmissible, as previously explained.

4.10 We saw in Part 3 that a forensic DNA profile consists of a set of individual locus profiles. An overall, composite likelihood ratio for a DNA profile is obtained by multiplying the likelihood ratios calculated for each individual locus in the chosen multiplex.

To simplify matters, we can begin by confining our attention to calculating the likelihood ratio for a single locus, on the assumption that the defendant's genotype matches the questioned profile depicted in Figure 4.1:



**Figure 4.1 EPG of questioned profile 16,17**

- 4.11 In practice, the conscientious forensic scientist would first satisfy herself that nothing has gone wrong anywhere in the profiling process and that the results obtained are valid. The heterozygote balance (explained in 3.11, above) would then be considered to ensure that it falls within the expected range of values for the profiling system in operation. Otherwise, the forensic scientist would need to review this part of the EPG in the light of the profile as a whole.
- 4.12 The procedure for calculating the likelihood ratio then begins by identifying a pair of propositions relevant to the case. Ideally, the forensic scientist should take into account whatever is known at this stage about the case the defence intends to run at trial. *Practitioner Guide No 4* will revisit these important practical issues. For present purposes, we can stick with the pair of sub-source propositions introduced as a standard illustration in 4.3, above:

***Prosecution Proposition*** (PP): the DNA recovered from the crime scene came from the accused.

***Defence Proposition*** (DP): the DNA recovered from the crime scene came from an unknown person, not blood related to the accused.

The likelihood ratio is the ratio of the probability of some evidence, E, assuming PP; to the probability of E, assuming DP. In mathematical notation,  $p(E | PP) / p(E | DP)$ .

In the context of DNA profiling, the probability of E assuming PP is the probability of a matching profile, assuming that the accused is the donor of the crime stain DNA. E is actually the conjunction of two ‘events’ (states of affairs): first, the probability of the crime stain profile (call it E<sub>1</sub>); secondly, the probability of the accused’s reference profile (call it E<sub>2</sub>). By applying the laws of probability, the likelihood ratio can be expressed as  $p(E_1 | E_2, PP) / p(E_1 | E_2, DP)$ .<sup>22</sup>

4.13 We will now illustrate this general approach to calculating likelihood ratios using the locus profile 16,17 depicted in Figure 4.1. The pair of probabilities to be fed into the likelihood ratio are:

$p(E_1 | E_2, PP)$ : the probability of obtaining the questioned profile 16,17, with peak heights 500 and 450 respectively, on the assumption that the prosecution proposition is true (the DNA comes from the accused).

$p(E_1 | E_2, DP)$ : the probability of obtaining the questioned profile 16,17, with peak heights 500 and 450 respectively, on the assumption that the defence proposition is true (the DNA comes from some other person, not blood related to the accused).

4.14  $p(E_1 | E_2, PP)$  is the probability of obtaining the questioned profile 16,17, if the donor has that genotype.

$p(E_1 | E_2, DP)$  is the product of two probabilities. First, the probability of obtaining the questioned profile 16,17, if the donor has that genotype. If we represent the unknown

---

<sup>22</sup> Some readers may be assisted by the following, somewhat simplified, formal proof. Let E<sub>1</sub> be the crime stain DNA profile, and E<sub>2</sub> the accused’s reference DNA profile. Let E be the evidence that E<sub>1</sub> and E<sub>2</sub> match, denoted (E<sub>1</sub>, E<sub>2</sub>). Then,

$$\begin{aligned} p(E | PP) / p(E | DP) &= p(E_1, E_2 | PP) / p(E_1, E_2 | DP) \\ &= \{p(E_1 | E_2, PP) p(E_2 | PP)\} / \{p(E_1 | E_2, DP) p(E_2 | DP)\} \end{aligned}$$

The DNA profile of the accused (E<sub>2</sub>) is independent of whether he was the donor of the DNA at the crime scene (PP) or some other person, not blood related to the accused, was (DP).

Thus  $p(E_2 | PP) = p(E_2 | DP)$  and these terms cancel out, leaving

$$p(E | PP) / p(E | DP) = p(E_1 | E_2, PP) / p(E_1 | E_2, DP).$$



donor as ‘U’, this probability can be expressed as  $p(E_1 | U, DP)$ . The second probability is the probability of someone other than the accused – an unknown person chosen at random from the relevant population – having genotype 16,17 at the specified locus, or  $p(U | E_2, DP)$ .

Notice that  $p(E_1 | E_2, PP)$  is exactly the same probability as  $p(E_1 | U, DP)$ , the probability of obtaining the questioned profile 16,17 if the donor is 16,17. Whatever value these probabilities might take, they cancel each other out and can be set aside (as demonstrated in n.22 on the facing page).

This simplifies the likelihood ratio to:

$$\text{Likelihood ratio} = \frac{1}{p(U | E_2, DP)}$$

(the probability of finding an unknown person at random in a relevant population whose genotype is 16,17, knowing the accused also has genotype 16,17)

4.15  $p(U | E_2, DP)$  has been dubbed the *random match probability* (or ‘random occurrence ratio’ by the Court of Appeal: see para.7.4, below). Although this terminology is sanctified by widespread usage, the whole concept of ‘a match’ becomes problematic as soon as we move away from single person profiles to consider mixed profiles or profiles generated from very small amounts of DNA. Conceptualising the issue in terms of random match probabilities can be helpful when calculating likelihood ratios for single person profiles. However, this conventional terminology, and the assumptions on which it rests, may need to be rethought if the types of evidence discussed in Parts 5 and 6 become more familiar features of criminal trials.

In order to calculate  $p(U | E_2, DP)$  it is necessary to ascertain the genetic composition of the relevant population, with further allowances for co-ancestry, sampling, and blood relatives.

4.16 *Allele frequencies*

The frequency of genotypes in a particular population is an empirical (biological) question. In theory, the DNA of everybody in the world could be profiled and a universal database of profiles constructed. Quite apart from the security, diplomatic and human rights implications of such an undertaking, this is unlikely to be a feasible political project for the foreseeable future.

In practice, the frequency of genotypes for the UK's major ethnic populations have been estimated from a database of donated samples.

4.17 Ethnicity is a factor in genetic variation. According to the 2001 Census, the population of England and Wales then comprised the following major ethnic groups:

**Figure 4.2 Population of England and Wales classified by ethnic appearance (2001 Census data)**

<b>Population</b> (Ethnic appearance group)	<b>Number of people</b>	<b>Percentage of People</b>
1. Caucasian	47,520,866	91.3
2. Asian of Indo-Pakistani origin	2,273,737	4.4
3. Black	1,139,577	2.2
4. Chinese or other	446,702	0.9
5. Mixed	661,034	1.3
Total	52,041,916	100.0

In many cases, the ethnic group of an unknown DNA donor will not be known. In other situations there may be eyewitness accounts indicating the perpetrator's apparent ethnicity. Here, the relevant variable is ethnic *appearance*: an eyewitness will generally be able to recount only the perpetrator's apparent ethnicity, inferred from that person's appearance, rather than the actual ethnic group, or groups, to which the perpetrator belongs.

4.18 Figure 4.3 reproduces statistical data collected on allele frequencies for one DNA locus, by ethnic sub-group:

**Figure 4.3 Allele counts for locus D3**

Allele	Population		
	Caucasian	Afro-Caribbean	Asian of Indo-Pakistani origin
9	0	0	0
11	0	0	0
12	1	0	0
13	5	1	1
14	115	26	24
15	232	88	110
15.2	0	0	0
16	216	115	125
17	170	76	88
18	123	21	47
19	12	1	5
20	0	0	0
<b>Total</b>	<b>874</b>	<b>328</b>	<b>400</b>

This classification, predicated on recent shared genetic inheritance, does not map precisely onto census data. In particular, the category ‘Afro-Caribbean’ is more selective than ‘Black’, which also covers direct immigration from Africa.

By extrapolating from these statistical data, it is possible to estimate the *allele probability* for each allele at that locus across the entire population. This is estimated by the *relative frequency* of the allele in the sample: i.e. the number of times that that particular allele occurred in the sample divided by the total number of all alleles obtained at that locus in the sample.

For example, the estimated probability of allele 16 in locus D3 in the Caucasian population is  $216/874 = 0.247$  (24.7%). This happens to be the second most common allele at D3 for sampled Caucasians. The estimated probability of allele 13 at locus D3, by contrast, is only  $5/874 = 0.057$  (5.7%) for Caucasians.

For the other two ethnic groups sampled, allele 16 in locus D3 is the most common allele at that locus: for Afro-Caribbeans,  $115/328 = 0.35$  (35%); and for Asians,  $125/400 = 0.313$  (31.3%).

4.19 It is a fair question whether a statistical database comprising only 1,600 allele counts is sufficient to allow one to estimate allele probabilities for the entire UK population of some 80 million people contributing up to 160 million alleles for each locus.

Generally speaking, the more counts there are in the database, the more precise is the relative frequency as an estimate of the allele probability.

However, the essential virtue of a sample is that it is adequate to support the particular kinds of statistical inference for which it is being used. Large samples are not necessarily required (leaving aside the question whether 1,600 counts *is* a large or small sample in absolute terms). DNA profiling is exceptionally discriminating between individuals, because it involves multiplying allele probabilities within each locus, and then further multiplying across the likelihood ratios for all ten or more loci in the multiplex. This procedure produces very robust composite likelihood ratios which, for full profiles, exceed one billion for unrelated individuals.

4.20 At the allelic level, the question is whether several hundred allele counts are sufficient for calculating ethnically-stratified allele probabilities. With appropriate values for allele probabilities, it can be shown that no more than several hundred alleles are required to generate robust estimates of allele frequencies when genotype probabilities are calculated using sampling and co-ancestry allowances such as those illustrated below. However, the adequacy both of the sampling allowance method and of the number of allele counts should always have been formally assessed using a statistical method like the one reported by Curran and Buckleton (2011).

#### 4.21 *Sampling allowance*

Allele counts with '0' in the database, such as allele 12 in locus D3 for non-Caucasians, present a problem. The zero count is assumed to be an artefact of sampling: otherwise, allele 12 at locus D3 would be treated as, in itself, uniquely identifying for any non-Caucasian with that profiled genotype.

However, simply increasing the size of the statistical database would not solve the problem. It is always possible to find new alleles, owing to the occurrence of genetic mutations.

4.22 Instead, forensic scientists apply a *sampling allowance* (known in the statistical literature as 'size bias correction') to compensate for the limited size of the database. Various methods of sampling correction are in use in different legal jurisdictions. One

standard approach involves adding the alleles actually observed in the case – both the suspect’s genotype and the questioned profile – to the database relative frequencies (Balding and Nichols, 1994). In effect, the case profiles are treated as additional empirical samples augmenting the existing database.

For example, taking the allele values 16,17 illustrated by Figure 4.1 for both reference and questioned profiles, the probability of allele 16 would be calculated as:

$$\text{for Caucasians:} \quad \frac{216 + 2}{874 + 4} = \frac{218}{878} = 0.248$$

and

$$\text{for Afro-Caribbeans:} \quad \frac{115 + 2}{328 + 4} = \frac{117}{332} = 0.352$$

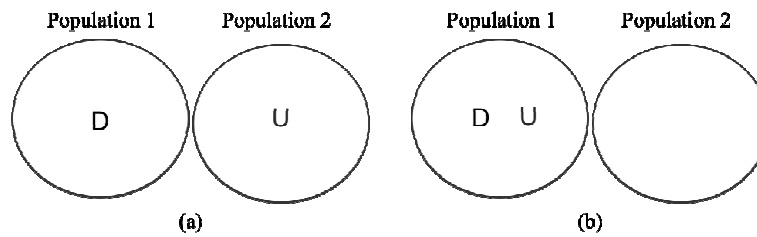
Notice that both accused and perpetrator each contribute two additional alleles, making four additional alleles in total, two of which are 16 and the other two are 17.

#### 4.23 *Co-ancestry allowance*

Since DNA is inherited from common ancestors, it is expected that the frequency of profiles will vary across ethnic groups. Some nationalities are more ethnically homogenous than others, but even in ethnically diverse countries like the UK many people will share common ancestors somewhere along the ancestral line. Moreover, people tend to intermarry within smaller groups for geographical, religious and cultural reasons. So, two people within an ethnic group are more likely to have a similar genotype than two people from different ethnic groups.

The impact of population genetics on the calculation of genotype probabilities and likelihood ratios is addressed through a *co-ancestry allowance*.

4.24 In any criminal trial, the accused’s ethnic appearance will obviously be known but (observing the presumption of innocence) the perpetrator’s ethnic appearance might not be. There are two logical possibilities, represented in Figure 4.4: either the defendant (D) and the unknown perpetrator (U) are from the same ethnic group or ‘population’, or they are from different populations:



**Figure 4.4 Unknown perpetrator U and defendant D are either from (a) different populations or (b) from the same population**

Intuitively, if D and U are from the same population, the probability of a shared genotype is somewhat greater because the incidence of that particular genetic characteristic (say, allele values 16,17 at a particular locus) may be greater in D and U's shared ethnic sub-population than in the general population. In these circumstances, the probative value of a matching profile will be correspondingly reduced. In mathematical terms, the likelihood ratio would be smaller because the genotype probabilities, forming the denominator of the likelihood ratio, would be larger.

4.25 The impact of possible co-ancestry on the likelihood ratio is accommodated through what is known as a *co-ancestry co-efficient*, sometimes also described as a *population sub-structure correction*, represented symbolically as  $F_{ST}$  or  $\theta$  (Balding and Nichols, 1994).

4.26 There are basically two ways of dealing with the situation in which U's ethnic appearance is entirely unknown.

First, one might simply assume that U and D are from the same ethnic population. This is a default assumption which may be justified, in the absence of any better information, because it favours the accused (reducing the size of the likelihood ratio indicates the diminished probative value of a matching profile), in accordance with the presumption of innocence.

A second alternative is known as *the stratification method* (Triggs et al., 2000). This involves modelling the ethnic composition of a relevant geographical region, calculating a genotype probability for each ethnic group in turn, and then taking the

weighted average of these genotype probabilities – weighted to reflect the ethnic mix in that region.

The stratification method aspires to be more systematic and objective than resorting to a default assumption. However, it does depend on having reliable statistical data on the ethnic compositions of particular regions, and moreover, requires the forensic scientist to make informed judgements about the relevant geographical population of potential perpetrators. Such judgements turn on circumstantial information about the nature of the crime and its commission (e.g. Did the perpetrator escape on foot or by car? Are we dealing with a domestic burglary or a contract killing by international drug smugglers? etc.).

#### 4.27 ***Blood relatives***

Further considerations need to be taken into account when calculating likelihood ratios for genotypes where the individuals concerned may be blood relatives. Consider, for example, the scenario where the accused claims that his brother is the guilty party (Evetts, 1992), or where suspicion falls on members of an extended family all residing in the same geographical region.

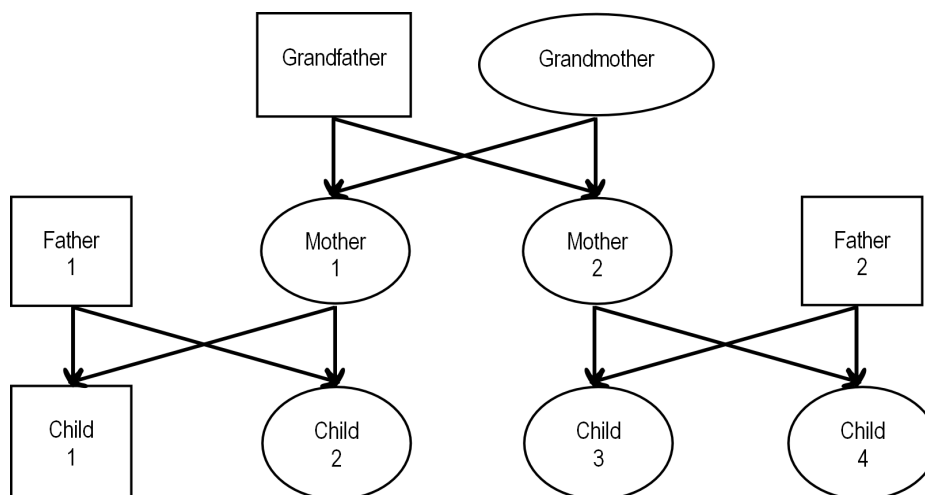
If reference samples can be obtained from all the relevant individuals, full DNA profiles can be generated precluding the need for any further statistical modelling. However, if particular (blood) relatives are uncooperative, unavailable or unknown, likelihood ratios must be constructed in the normal way in order to assess the probative value of DNA profiling evidence. In order to calculate a likelihood ratio the forensic scientist would need to postulate potential defence propositions based on known information, such as the proposition ‘the source of the DNA is the accused’s brother’. In addition to allele probabilities, sampling allowance, and the co-ancestry coefficient, that is to say, the forensic scientist must now also take account of possible blood-relatedness.

4.28 The probative value of DNA evidence is reduced if there is a realistic possibility that the perpetrator and the defendant are blood relatives. Whilst members of all genetically-linked populations share distant co-ancestry, blood relatives have much more proximate common ancestors in recent generations. The more recent this shared

ancestry, the smaller the probative value of DNA profiles in distinguishing between blood relatives.

4.29 When we calculated likelihood ratios for unrelated individuals earlier in this Part, we estimated the probability of their having shared alleles purely by chance. This is the genotype probability based on allele counts. For blood relatives, however, the salient factor becomes recent genetic inheritance – recalling that alleles are inherited from both parents, who inherited theirs from their parents, and so on, back up the line of genetic descent.

For example, siblings are more likely to share inherited alleles from their parents than are first cousins, as Figure 4.5 illustrates diagrammatically:



**Figure 4.5 Family tree of siblings and first cousins**

In Figure 4.5 males are shown by squares, and females by circles. Each arrow represents one allele donated by a parent. Mendel’s theory of heredity posits that a parent is equally likely to pass on either of their two alleles to their offspring. Consequently, the probability that a particular allele will be passed on to a particular child is halved in each successive generation.

4.30 Figure 4.6 lists the types of blood relationship most frequently encountered in forensic casework, with the percentage of pairs of individuals sharing none, one or two



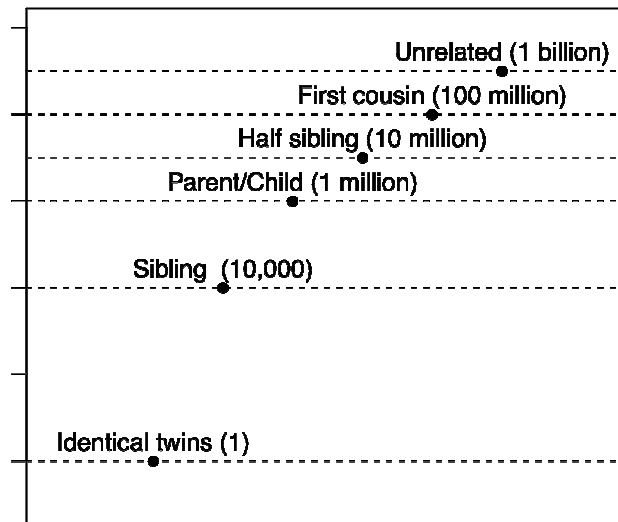
inherited alleles at any locus. Identical twins and non-blood relatives are included for purposes of comparison.

**Figure 4.6 Inherited alleles by relationship**

Blood relationship	Percentage of two people sharing		
	<i>None</i>	<i>One</i>	<i>Two</i>
1. Identical twins			100
2. Parent/child		100	
3. Siblings	25	50	25
4. Half-siblings	50	50	
5. Uncle/nephew	50	50	
6. Grandparent/grandchild	50	50	
7. First cousins	75	25	
8. Unrelated	100		

At one end of the continuum, identical twins share both alleles by direct inheritance; at the other pole, people that are not blood related have no directly inherited alleles in common. Siblings are more likely than first cousins to have the same alleles, because they have inherited all of their alleles from their common parents. However, we can see from Figure 4.6 that siblings are just as likely to have no alleles in common as to share both alleles (there is a 0.25 or 25% probability of either eventuality). By contrast, there is a 0.75 (or 75%) probability that first cousins would share no inherited alleles and a 0.25 (or 25%) probability that they have one shared allele. These simple calculations demonstrate why genetic proximity reduces the size of the likelihood ratio and correspondingly decreases the probative value of a matching DNA profile.

4.31 Foreman and Evett (2001) calculated likelihood ratios across loci for the most common genotypes in the SGMPlus multiplex for several possible blood relationships between the accused and the perpetrator. The results of their analysis are summarised by Figure 4.7:



**Figure 4.7 Likelihood ratios for possible blood relationships (rounded by orders of magnitude in multiples of 10)**

At one extreme, the likelihood ratio for identical twins is one – both twins have the same genotype, and DNA profiling cannot distinguish between them. At the other extreme, the likelihood ratio for unrelated, non-blood relatives is one billion. This simple contrast graphically illustrates the pronounced impact of relatedness on likelihood ratios (i.e. on the probative value of DNA evidence).

4.32 The order of magnitude rapidly increases as blood relationships become more distant. This is because, as we have seen, a full DNA profile (genotype) comprises multiple loci, and the likelihood ratio for a full matching profile is calculated by multiplying across the allele probabilities for each locus (with co-ancestry allowance). The probabilities of matching full DNA profiles soon become very small, even for blood relatives.

4.33 Combining the results shown in Figures 4.6 and 4.7, we can see that the likelihood ratios for uncle/nephew (or aunt/niece, etc) and grandparent/grandchild relationships will be of the same order of magnitude as the likelihood ratio for half-siblings. In each of these relationships, there is a 0.5 probability, or 50% chance, of one shared allele and a 0.5 probability/50% chance of no shared alleles.

The probability that a first cousin of the accused will share his genotype is of the order of 100 million. There is no need to consider more distant blood relationships, such as second cousins etc, because in these cases the likelihood ratio increases to the

order of one billion – i.e. as large as the likelihood ratio for unrelated individuals. This shows that, for the purpose of constructing likelihood ratios in DNA profiling, relatives more distant than first cousins can be treated in exactly the same way as any other person in the relevant population, applying the co-ancestry allowance.

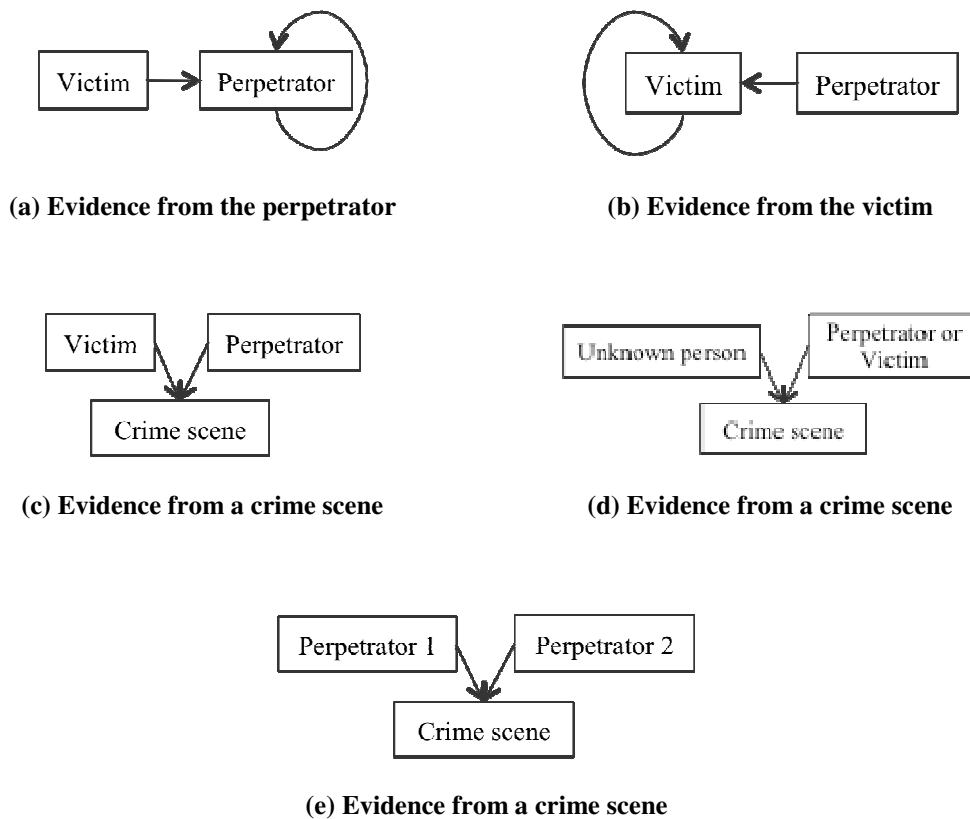
## 5. Two Person ('Mixed') Questioned Profiles

5.1 Part 4 explained how to calculate likelihood ratios for a full one-person profile. This Part addresses the further issues that must be tackled when the forensic scientist is confronted with mixed, two-person profiles. However, we are still making the simplifying assumption that there is an optimal amount of DNA from which to construct full profiles.

### 5.2 *Transfer patterns*

In situations where there is only one potential donor of DNA, transfers of genetic material may occur from the perpetrator to the victim or to the scene, or from the victim to the perpetrator or to the scene. Possible routes of primary transfer obviously multiply in scenarios involving two or more potential donors, as indicated by Figure 5.1:

Figure 5.1 Transfer patterns for two person questioned profiles



In situations (a) and (b) depicted in Figure 5.1, DNA from an incident is deposited onto background DNA from an existing source. An example of (a) is where blood from the victim is deposited onto the perpetrator's clothing, which will already contain traces of the perpetrator's own DNA from normal wear. A blood sample from this source might then produce a mixed profile comprising DNA contributed by both victim and perpetrator. Similarly, if an intimate swab is taken from a rape victim, as modelled in scenario (b), DNA from the semen of the rapist might be mixed with DNA from the vaginal cells of the victim. Although it is possible to isolate and extract DNA from semen, this procedure fails to remove all of the victim's DNA in about one third of cases. Scenario (b) is confronted whenever a stain is a mixture of body fluids that cannot be physically disaggregated.

Situation (c) might arise, for example, when the victim has been hit with a hammer which is later recovered from the scene. The handle of the hammer may contain skin cells from the perpetrator as well as from the victim. The question to be addressed is whether the victim and the suspect are in fact the two donors of the mixed profile, or whether some third party might be the source of the DNA (possibly indicating that unknown person's involvement in the incident).

Situation (d) is exemplified when the handle of a tool used in housebreaking contains DNA from the perpetrator as well as from an unknown donor. The unknown person could be the owner of the property or a lawful visitor, or someone who had previously handled the tool quite innocently. Situation (e) arises when, for example, accomplices both handle an implement used in the crime (e.g. a jemmy) or share clothing, such as balaclava masks.

5.3 The five 'situations' just identified can be reduced to three general scenarios commonly encountered in forensic case-work. In each scenario, a mixed profile contains the accused's or the victim's DNA and that of one other person:

**Scenario #1: a second known donor whose identity is not in dispute in the proceedings:** e.g. a vaginal swab contains DNA from the complainant and an identified alleged assailant. The DNA profile can be assessed in light of the genotype

of the known second donor. Transfer patterns (a) and (b), and also occasionally (c), fall within this scenario.

**Scenario #2: a second unknown donor whose identity is not in dispute in the proceedings:** e.g. a sample taken from the mouth area of a balaclava left at the scene of a crime often contains profiles from more than one person. The donors might be the alleged perpetrator and a previous wearer of the balaclava whose identity is unknown. This is an illustration of transfer pattern (d) in Figure 5.1.

**Scenario #3: a second suspected donor whose identity is disputed in the proceedings:** e.g. co-defendants (D1 and D2) pleading not guilty, as in situation (e). The default assumption made in relation to each defendant in turn is that the second donor is unknown, allowing the forensic scientist to adopt essentially the same approach as in Scenario #2 cases. This is equivalent to assuming, for the purposes of DNA profiling, that D1 and D2 would be tried separately in different trials; and it involves no assumption inconsistent with the presumption of innocence.

Scenario #3 also embraces cases potentially involving two-way transfers, as depicted in situation (c) in Figure 5.1. For example, where a mixture potentially containing the DNA of both victim and perpetrator is recovered from a knife of unknown provenance, the identity of both donors may be disputed (because, e.g., the defence does not accept that the knife was used in the alleged assault).

5.4 It is possible to construct likelihood ratios for each of these three scenarios, adapting the probabilistic methodology introduced in the previous Part.

We will now present four detailed illustrations of how to calculate likelihood ratios for different kinds of mixed profiles. The first three illustrations all relate to Scenario #1 type cases, and involve, respectively: (a) a balanced four peak profile; (b) a three peak profile; and (c) an unbalanced three peak profile. The fourth illustration, (d), relates to a balanced four peak profile for Scenario #2 (which, as we have just seen, also covers Scenario #3, as well). The techniques elucidated in (b) and (c) could easily be extrapolated to Scenario #2 situations, but we have not undertaken that lengthy and somewhat repetitive exercise here. Between them, our four illustrations demonstrate

how to tackle most of the issues that might be confronted in deriving, or interpreting, likelihood ratios for mixed DNA profiles.

It bears repeating that appropriate scenario selection, in conjunction with the formulation of a relevant pair of competing propositions, must first be accomplished by applying logical analysis to the circumstances of the instant case, as the forensic scientist understands them to be, before any computation of the likelihood ratio can be attempted.

5.5 (a) *Questioned Profile with four balanced peaks for Scenario #1*

Suppose that DNA profiling produces the following ‘balanced’ questioned profile, with four peaks – indicating two donors, each donating two of the four alleles:

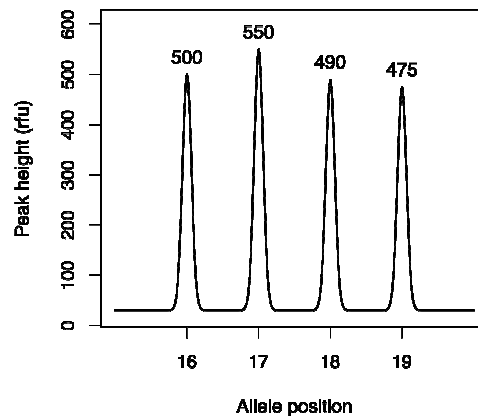


Figure 5.2 A balanced questioned profile from two people

In Scenario #1, the second donor is the complainant whose identity is known and undisputed. Here, the accused’s genotype for this locus (D3) is 16,17, and the complainant’s genotype is 18,19 (with peak heights 500, 550, 490, and 475, respectively).

5.6 The relevant competing propositions are:

**Prosecution proposition (PP):** the DNA recovered from the questioned sample came from the complainant and the defendant; and

**Defence proposition (DP):** the DNA recovered from the questioned sample came from the complainant and an unknown person unrelated to the defendant and the complainant.

The likelihood ratio is the ratio of the two probabilities:

$p(E | PP)$ : the probability of the DNA evidence, E, on the assumption that the prosecution proposition is true; and

$p(E | DP)$  : the probability of the DNA evidence, E, on the assumption that the defence proposition is true.

In this example, the evidence, E, is the conjunction of three ‘events’ (states of affairs): the questioned profile ( $E_1$ ), the defendant’s reference profile ( $E_2$ ) and the complainant’s reference profile ( $E_3$ ). Applying the laws of probability (exemplified by n.22 to para.4.12, above), the likelihood ratio is the ratio of the two probabilities,  $p(E_1 | E_2, E_3, PP)$  and  $p(E_1 | E_2, E_3, DP)$ .

- 5.7  $p(E_1 | E_2, E_3, PP)$  is the probability of obtaining the questioned profile given that:
- (a) the defendant’s genotype is 16,17;
  - (b) the complainant’s genotype is 18,19; and
  - (c) the defendant and the complainant are the donors of the questioned profile (the prosecution proposition).

In other words, the probability forming the numerator of the likelihood ratio is:

$p(E_1 | E_2, E_3, PP) =$  *the probability of obtaining the questioned profile given that the genotypes of the donors are 16,17 and 18,19*

- 5.8  $p(E_1 | E_2, E_3, DP)$  is the probability of obtaining the questioned profile given that:
- (a) the defendant’s genotype is 16,17;
  - (b) the complainant’s genotype is 18,19; and
  - (c) the complainant and an unknown person, unrelated to the defendant or the complainant, are the donors of the questioned profile (the defence proposition).



5.9 The defence proposition nominates an unknown person, together with the complainant, as donors of the questioned profile. This nominated individual comes from a *population* of unknown people. However, not all the people in the population could be donors of the profile. In fact, only a person with genotype 16,17 could be the donor of the profile, in view of the fact that alleles 18,19 are already accounted for – by the complainant.

It follows that the denominator of the likelihood ratio is the product of two probabilities:

$$p(E_1 | E_2, E_3, DP) = \frac{p1: \text{the probability of obtaining the questioned profile given that the genotypes of the donors are 16,17 and 18,19}}{\times p2: \text{the probability of finding a person with genotype 16,17 in a population}}$$

The numerator probability,  $p(E_1 | E_2, E_3, PP)$ , and the probability in the first factor of the denominator ( $p1$ ) are the same. Their ratio is one and they can be removed (because they ‘cancel out’) in the final calculation, as previously explained in relation to single person profiles.

The likelihood ratio of  $p(E | PP)$  to  $p(E | DP)$  is then given by:

$$\text{Likelihood ratio} = \frac{1}{p2: \text{the probability of finding a person in a relevant population whose genotype is 16,17}}$$

5.10 Probability  $p2$  is calculated using the allele-count databases and probabilistic methods described in Part 4, together with the known genotypes of the defendant and the complainant.

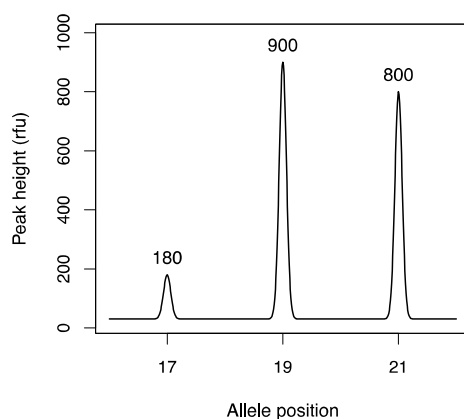
If both complainant and defendant are from the same population as the potential unknown donor, the genotype probability would be adjusted slightly to take account of the additional information available to us regarding the alleles of two more people in the population – the accused and the complainant. Since the accused has genotype 16, 17, the probability of another person in that population having genotype 16, 17 is increased by a tiny amount. But the probability also decreases by a tiny amount, to take account of the fact that a second member of the population, the complainant,

does not have genotype 16,17 at that locus. This is another illustration (like the sampling allowance technique explained in paragraphs 4.21-4.22, above) of the general approach described in *Practitioner Guide No 1* to up-dating conditional probabilities in the light of new information.

5.11 **(b) Questioned Profile with three peaks for Scenario #1**

Two person profiles do not always show up as four peaks at a particular locus, in the way illustrated by Figure 5.2. Most obviously, if both donors share the same genotype at that locus only two peaks will be shown on the EPG.

Sometimes, profiles exhibit three peaks, as illustrated by Figure 5.3:



**Figure 5.3 Questioned profile from two donors with three peaks**

For Scenario #1 situations, the profiles of both potential donors are known. Let us suppose that the accused's genotype is 17,19 and the complainant's genotype is 19,21 at this locus. It is possible that in the profile depicted by Figure 5.3 the complainant's peak at allele 19 is *masking* the accused's, and this can be factored into the calculation of the likelihood ratio.

5.12 The competing propositions are the same as in the previous example:

**Prosecution proposition (PP):** the DNA recovered from the questioned sample came from the complainant and the defendant; and

**Defence proposition (DP):** the DNA recovered from the questioned sample came from the complainant and an unknown person unrelated to the defendant and the complainant.

Applying the same logic as before, the numerator of the likelihood ratio is given by:

$$p(E_1 | E_2, E_3, PP) = \text{the probability of obtaining the questioned profile given that the donors' genotypes are 17,19 and 19,21.}$$

5.13 The defence proposition postulates an unknown person as a co-donor of the mixed profile with the complainant. This unknown person could logically have any one of three possible genotypes: 17,17; 17,19; or 17,21. The unknown person must have contributed allele 17, because we know that this allele did not come from the complainant.

The probability of any one of three mutually exclusive 'events' is calculated by adding up their individual probabilities (just as the probability of rolling a 1, 2 or 3 on a six-sided die is the probability of rolling 1, plus the probability of rolling 2, plus the probability of rolling 3 =  $1/6 + 1/6 + 1/6 = 3/6 = 1/2$ ).

Likewise, we can calculate  $p(E_1 | E_2, E_3, DP)$ , the probability of the evidence assuming that the defence proposition is true (supplying the denominator of the likelihood ratio), by adding together the probability of each candidate genotype in the population, 17, 17; 17, 19; and 17, 21:

$$\begin{aligned}
 p(E_1 | E_2, E_3, DP) = & \text{p1: the probability of obtaining the questioned profile given that} \\
 & \text{the donors' genotypes are 17,17 and 19,21} \\
 & \times \\
 & \text{p2: the probability of finding a person in a population whose} \\
 & \text{genotype is 17,17} \\
 & + \\
 & \text{p3: the probability of obtaining the questioned profile given that} \\
 & \text{the donors' genotypes are 17,19 and 19,21} \\
 & \times \\
 & \text{p4: the probability of finding a person in a population whose} \\
 & \text{genotype is 17,19} \\
 & +
 \end{aligned}$$

p5: the probability of obtaining the questioned profile given that the donor's genotypes are 17,21 and 19,21

×

p6: the probability of finding a person in a population whose genotype is 17,21

5.14 The genotype probabilities p2, p4 and p6 are calculated by considering ethnic appearance population, allele databases, sampling, and co-ancestry allowances, as explained in Part 4.

The probabilities of obtaining the questioned profile given the genotypes of the putative donors (p(E | PP), p1, p3 and p5) are assessed by reference to heterozygote balance (see paras.3.11-3.14, above) and the **mixing proportion** across loci, which is explained in the next paragraph. The forensic scientist satisfies herself that each of the probabilities p1, p3 and p5 is greater than zero, using the method described below, so that they can be eliminated from the calculation of the likelihood ratio by cancelling through, as before.

If p(E | PP) = 1, as in previous illustrations, then the likelihood ratio is given by:

$$\text{Likelihood ratio} = \frac{1}{\begin{array}{l} \text{p2: the probability of finding a person in a population whose} \\ \text{genotype is 17,17} \\ + \\ \text{p4: the probability of finding a person in a population whose} \\ \text{genotype is 17,19} \\ + \\ \text{p6: the probability of finding a person in a population whose} \\ \text{genotype is 17,21} \end{array}}$$

Notice, once again, that the forensic scientist's professional judgement and expertise necessarily enter into assessments of whether probabilities p1, p3 and p5 are approximately the same, and thus cancel out.

5.15 *Mixing proportion*

Calculations of likelihood ratios for two person mixed profiles employing the procedure described in the previous paragraphs are subject to making appropriate allowance for the **mixing proportion**. This is the proportion of DNA contributed by a

particular donor to the profiled mixed sample. The mixing proportion is expressed as a number between zero and one, e.g. a mixing proportion of 0.7 means that one donor has contributed 70% of the DNA while the other donor has contributed the remaining 30%.

The mixing proportion is used in conjunction with measurements of heterozygote balance, preferably through the application of computerised statistical models. Peak heights are split according to the mixing proportion and donor genotypes, facilitating consideration of heterozygote balance between the assigned peak heights. If the resulting peak balance falls outside the expected range, the probability of obtaining the questioned profile given the putative donors is deemed to be zero. In other words, these candidates are eliminated as potential donors of the DNA in the mixed profile.

5.16 There are various ways to calculate an estimate of the mixing proportion by considering the peak heights across all loci.

Consider the following questioned profile with four peaks at a locus:

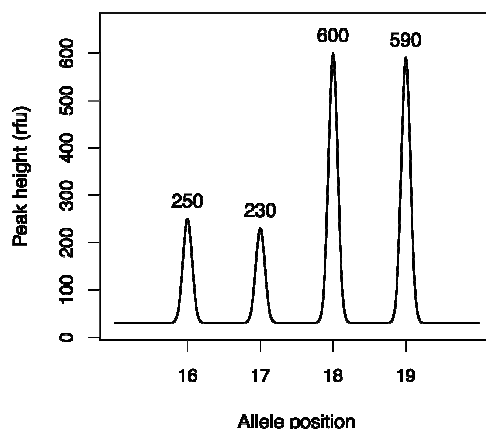


Figure 5.4 A balanced questioned profile with two donors

In this questioned profile, heterozygote balance suggests that peaks 16 and 17 form a heterozygous pair, as do peaks 18 and 19. The mixing proportion is calculated by adding the peak heights of a pair and dividing it by the sum of the heights of all the peaks in the profile, as shown in Figure 5.5.

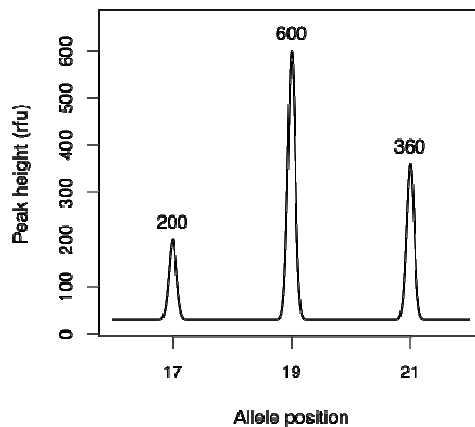
**Figure 5.5 The mixing proportion of the profile in Fig 5.4**

	Donor 1		Donor 2	
Genotype	16	17	18	19
Height	250	230	600	590
Height sum per donor	480		1190	
Mixing proportion	480/1670 = 0.29		1190/1670 = 0.71	

In this example, Donor 1 contributed 29% of the DNA in the sample, and Donor 2 contributed 71%. This means that the probability of obtaining the questioned profile if the donors' genotypes are 16,17 and 18,19 is greater than zero.

5.17 (c) *Questioned Profile with three unbalanced peaks for Scenario #1*

Consider a further variation on a profiling result, illustrated by Figure 5.6:



**Figure 5.6 Questioned profile from two donors with three unbalanced peaks**

We are still assuming a Scenario #1 situation (two known donors), but the accused's genotype in this illustration is known to be 17, 19 and the complainant's genotype is 19, 21 at this locus. The additional factors that must now be considered, concurrently, are *masking* and *peak height imbalance*.

5.18 The likelihood ratio is calculated by applying exactly the same formula as in the previous two illustrations:

$$\text{Likelihood ratio} = \frac{1}{\begin{aligned} & p2: \text{the probability of finding a person in a population whose} \\ & \text{genotype is 17,17} \\ & + \\ & p4: \text{the probability of finding a person in a population whose} \\ & \text{genotype is 17,19} \\ & + \\ & p6: \text{the probability of finding a person in a population whose} \\ & \text{genotype is 17,21} \end{aligned}}$$

However, whereas the EPG profile in Figure 5.3 was capable of supporting all three possibilities represented by p2, p4 and p6, the three peak profile in Figure 5.6 is consistent only with p2 and p4, whilst ruling out p6. When heterozygote balance is taken into account, the probability of obtaining the profile shown in Figure 5.6 if the unknown donor has genotype 17,21 at that locus is shown to be close to zero.

5.19 Adapting the procedure introduced in Part 3, heterozygote balance can be calculated by ‘sharing’ the height of peak 21 between peaks 17 and 19 in a variety of combinations. For example, if we assigned 200 rfu of peak 21 to be paired with peak 17, the *heterozygote balance* for this pair would be 1 (since the height of peak 17 is also 200 rfu). But this would leave only 160 rfu to be paired with peak 19, with height 600. The *heterozygote balance* for this pair is 160/600 = 0.27, which, let us assume, is outside the expected range of values for the relevant profiling system. Figure 5.7 summarises this result:

	Unknown person		Complainant	
Genotype	17	21	19	21
Height	200	200	600	160
Heterozygote balance	1.00		0.27	

**Figure 5.7 Heterozygote balance for genotype 17, 21**

Every possible way in which peak 21 could be shared between peaks 17 and 19 can be systematically considered. If none of these variations produces a heterozygote balance within the expected range, then the probability of obtaining the questioned profile given donor genotypes 17,21 and 19,21 approximates to zero.

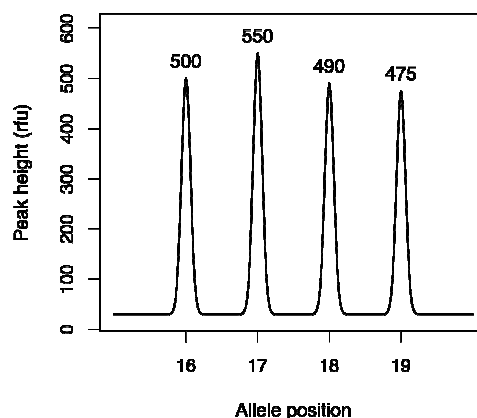
5.20 Having discounted 17,21 as a possible donor genotype, the likelihood ratio for the profile in Figure 5.6 becomes:

$$\text{Likelihood ratio} = \frac{1}{\begin{array}{l} p_2: \text{the probability of finding a person in a population whose} \\ \text{genotype is 17,17} \\ + \\ p_4: \text{the probability of finding a person in a population whose} \\ \text{genotype is 17,19} \end{array}}$$

This likelihood ratio will be larger than that for the profile depicted in Figure 5.3, which is intuitively correct. There is bound to be a greater probability of finding a person in a population with any one of the three genotypes 17,17 or 17,19 or 17,21 than of finding a person in the same population with (only) genotypes 17,17 or 17,19, provided that the probability of genotype 17,21 is greater than zero. As the probability of the denominator increases, the likelihood ratio becomes smaller. (If the denominator covered every possible allele pair at that locus, the likelihood ratio would shrink all the way down to  $1/1 = 1$ .)

5.21 **(d) Questioned Profile with four balanced peaks for Scenario #2**

Our final illustration replicates the scenario of a mixed profile where the identity of one of the donors is disputed and the unknown donor is not in dispute. To simplify matters somewhat, we will reuse the example of the balanced profile previously given in Figure 5.2:



There are four peaks at this locus, with heights 16, 17, 18 and 19. The profile of the accused is 16,17 at this locus.



5.22 The two competing propositions forming our likelihood ratio pair are:

**Prosecution proposition (PP):** the DNA recovered from the questioned sample came from the defendant and an unknown person unrelated to the defendant; and

**Defence proposition (DP):** the DNA recovered from the questioned sample came from two unknown people unrelated to the defendant or to each other

In this example, the evidence, E, is the conjunction of two ‘events’: the crime stain profile (E<sub>1</sub>) and the defendant’s reference profile (E<sub>2</sub>). Applying the laws of probability as before (see n.22, para.4.12), the likelihood ratio can be shown to be the ratio of the two probabilities, p(E<sub>1</sub>|E<sub>2</sub>,PP) and p(E<sub>1</sub>|E<sub>2</sub>,DP).

5.23 The numerator of the likelihood ratio for the evidence, E, is

$$p(E_1 | E_2, PP) = \text{the probability of obtaining the questioned profile given that the defendant's genotype is 16,17 and that the prosecution proposition is true}$$

The prosecution proposition nominates the defendant and an unknown person as the donors of the questioned profile. The numerator can be rewritten as:

$$p(E_1 | E_2, PP) = \begin{matrix} p1: \text{the probability of obtaining the questioned profile given that the} \\ \text{genotypes of the donors are 16,17 and 18,19} \\ \times \\ p2: \text{the probability of finding a person in a population whose genotype} \\ \text{is 18,19} \end{matrix}$$

In principle, any person in the population could be a donor, and the numerator of the likelihood ratio needs to take this into account. However, on the prosecution’s proposition, the probability of obtaining the questioned profile when the unknown donor has a genotype other than 18,19 is zero. This is because the prosecution’s proposition cannot be true unless the accused is the donor of alleles 16 and 17, which in turn implies that the unknown donor of the mixed sample must have contributed alleles 18 and 19. Hence, p2 is the only other probability relevant to calculating the numerator of the likelihood ratio.

5.24 The denominator of the likelihood ratio is:

$$p(E_1 | E_2, DP) = \begin{array}{l} \textit{the probability of obtaining the questioned profile given that the} \\ \textit{defendant's genotype is 16,17 and that the defence proposition} \\ \textit{is true} \end{array}$$

5.25 The defence proposition postulates two unknown donors of the questioned profile. There are six possible ways of separating the four alleles into pairs, i.e. six possible genotypes. The possible genotype pairs are set out in Figure 5.8:

Genotype pair number	Genotype 1	Genotype 2
1	16,17	18,19
2	16,18	17,19
3	16,19	17,18
4	17,18	16,19
5	17,19	16,18
6	18,19	16,17

**Figure 5.8 Six genotype pairs for the questioned mixed profile in Figure 5.2**

Factoring these possibilities into the calculation of the likelihood ratio, the denominator becomes:

$$p(E_1 | E_2, DP) = \begin{array}{l} p_3: \textit{the probability of obtaining the questioned profile given that the} \\ \textit{donors' genotypes are genotype pair 1} \\ \quad \times \\ p_4: \textit{the probability of finding two people in a population whose} \\ \textit{genotypes are genotype pair 1} \\ \quad + \\ \quad \cdot \\ \quad \cdot \\ \quad \cdot \\ \quad + \\ p_{13}: \textit{the probability of obtaining the questioned profile given that the} \\ \textit{donors' genotypes are genotype pair 6} \\ \quad \times \\ p_{14}: \textit{the probability of finding two people in a population whose} \\ \textit{genotypes are genotype pair 6} \end{array}$$

5.26 In formal terms, all of the six genotype pairs represent possible combinations of donors. However, some of these theoretical combinations could safely be eliminated for all practical purposes. When heterozygote balance and mixing proportions across

all loci were considered for each of the probabilities  $p_3$ ,  $p_5$ ,  $p_7$ ,  $p_9$ ,  $p_{11}$  and  $p_{13}$  some of them would approximate zero, and could be excluded from further analysis.

Calculation of the genotype probabilities ( $p_4$ ,  $p_6$ ,  $p_8$ ,  $p_{10}$ ,  $p_{12}$  and  $p_{14}$ ) would be performed utilising databases of allele counts, and making appropriate allowances for ethnic appearance, sampling and co-ancestry, in the routine manner described in Part 4. Although mixed profiles introduce additional complications, especially if the identity of one of the donors is unknown or disputed (as in Scenario #2 and #3 type-situations), the basic approach to calculating likelihood ratios, utilising a relatively small number of logical axioms, inferences and assumptions, remains the same.

## 6. Low Template DNA (LTDNA)

6.1 To this point, we have been assuming that crime stains and questioned samples contain sufficient genetic material to produce a DNA profile employing the standard STR profiling method described in Part 2. But this is not always the case. Crime stains may contain minute amounts of DNA, requiring special techniques to be applied in order to generate a profile. These techniques have further implications for making probabilistic assessments of the probative value of DNA evidence.

### 6.2 *From LCN to LTDNA*

In the past, the term *low-copy-number* (LCN) profile has been used to refer to profiles obtained from very small amounts of DNA (Gill et al., 2000). However, LCN profiles are closely associated with the use of 34 amplification cycles, instead of the standard 28 cycles.

More recently it has been recognised that, given the increased sensitivity of techniques for producing profiles, it is possible to obtain profiles from small amounts of DNA employing a variety of methods. The broader generic term ‘low-template DNA’ (Caddy et al., 2008), often abbreviated to LTDNA, was coined to describe any analytical process generating profiles from limited DNA template.

6.3 LCN attracted negative judicial comment from the Northern Ireland Crown Court in *R v Hoey*,<sup>23</sup> but the reliability of LTDNA profiles was subsequently affirmed by the English Court of Appeal in *R v Reed and Reed*,<sup>24</sup> where the Court stated:

[A] challenge to the validity of the method of analysing Low Template DNA by the LCN process should no longer be permitted at trials where the quantity of DNA analysed is above the stochastic threshold of 100-200 picograms in the absence of new scientific evidence...<sup>25</sup>

On this interpretation, the *stochastic threshold* refers to the minimum amount of DNA needed to produce a reliable profile using the LTDNA method. The Court of

---

<sup>23</sup> *R v Sean Hoey* [2007] NICC 49 (20 December 2007), available from BAILII.

<sup>24</sup> *R v Reed and Reed; R v Garmson* [2010] 1 Cr App R 23; [2009] EWCA Crim 2698.

<sup>25</sup> *Ibid.* [74].

Appeal in *Reed* observed that ‘[t]here is no agreement among scientists as to the precise line where the stochastic threshold should be drawn, but it is between 100 and 200 picograms’.<sup>26</sup>

However, according to an alternative conception propounded, for example, by the FBI’s Scientific Working Group on DNA Analysis and Methods (SWGDM, 2010), the stochastic threshold is the rfu value (peak height) at which it is reasonable to assume that the ‘sister allele’ of a heterozygous pair has not suffered *allelic drop-out* (see para.6.8, below). When ‘stochastic threshold’ is understood in this way, the particular DNA system being employed is a relevant consideration in addition to the amount of DNA available for testing (Gill et al., 2009).

#### 6.4 *Practical benefits of LTDNA profiling*

Resort to the more sensitive LTDNA technique may be necessary in order to generate a usable DNA profile in the following types of scenario, which are regularly encountered in practice:

- DNA deposited at crime scenes might not be in the form of visible or chemically identifiable stains suitable for profiling systems designed to work with optimal amounts of DNA. For example, items handled by the perpetrator, such as a knife handle or jemmy, may retain loose skin cells containing small amounts of DNA.
- In some cases DNA has degraded because the crime stain has been exposed to heat or humidity in the environment. It may then not be possible to extract enough good quality DNA from the crime stain to satisfy the requirements of standard STR profiling.
- In other cases genetic material connected to a crime, e.g. blood in soil, cannot be separated from chemicals that inhibit the amount of DNA available for profiling.

---

<sup>26</sup> Ibid.

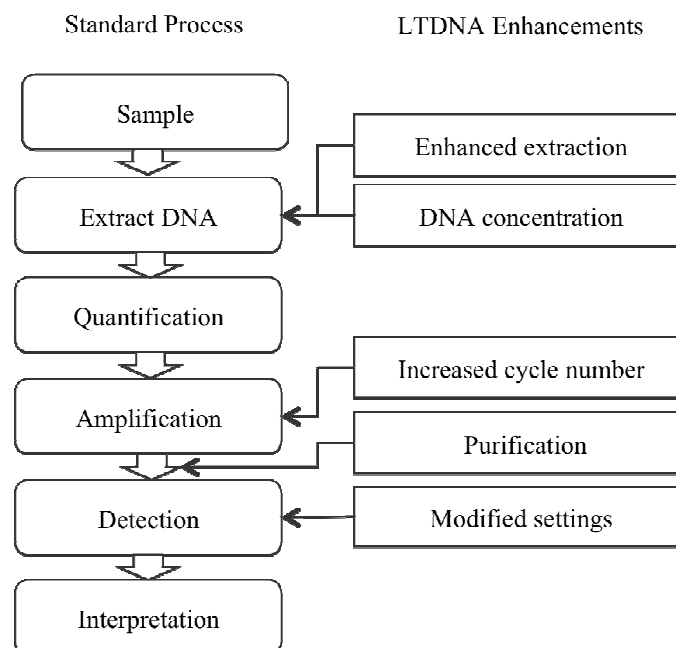
- Another potential difficulty is that one of the donors to a mixed sample may have contributed only a very small amount of the DNA it contains. This is a familiar problem in relation to sexual offences, in particular, where the perpetrator's contribution of DNA to a mixed profile may be very much smaller than the victim's contribution.

In all of these scenarios, the forensic potential of DNA profiling would be exhausted, but for the possibility of utilising more sensitive LTDNA techniques.

## 6.5 Profiling enhancements

We saw in Part 2 that a *DNA system* consists of a multiplex (including the chemicals used in DNA amplification and the specification of selected loci to profile) and a protocol (the equipment and settings used in the profiling process). The sensitivity of a DNA system can be enhanced at the design stage of the multiplex and/or by making adjustments to the protocol. The general idea is to obtain more DNA of better quality from a sample, and to amplify and detect still more of it (see Hopwood et al., 2011).

Figure 6.1 indicates, schematically, the points at which there is scope for making LTDNA enhancements to the standard profiling process:



**Figure 6.1 LTDNA enhancements to standard profiling processes**

6.6 A profile can be produced from an LTDNA sample using various profiling systems. The two systems most commonly employed in the UK at present are SGMPlus at 28 and at 34 amplification cycles. An LTDNA sample analysed with SGMPlus at 28 cycles may produce a profile with short peaks which are difficult to distinguish from background ‘noise’.

New multiplexes implementing the European Standard Set (ESS) of loci are more sensitive than the SGMPlus multiplex in current service. These new generation multiplexes employ 29 or 30 amplification cycles, and will be capable of producing profiles from LTDNA samples as standard (Tucker et al., 2011; Tucker et al., 2012).

### 6.7 *Enhanced anti-contamination procedures*

The increased sensitivity of LTDNA profiling means that enhanced anti-contamination measures are vital for maintaining the integrity, and potential probative value, of DNA profiles generated using this method. The following precautions should be considered essential:

- (i) Every physical location at which the profiling process takes place must be designed and maintained DNA free (e.g. ventilation systems should direct airflow outwards not inwards).
- (ii) Each stage in the process indicated by Figure 6.1 should be performed in a separate room. Staff and samples should flow through the profiling process in a single, predetermined direction in order to prevent contaminants, especially those arising from the amplification room, from infecting previous stages in the process.
- (iii) Laboratories should be tested regularly for the presence of contaminating DNA.
- (iv) An *elimination database* of scientific staff, consumable manufacturers and police should be maintained.<sup>27</sup> Should contamination ever be detected, the contaminant DNA profile can then be searched against the elimination database and the donor of the DNA, if identified, can be eliminated from

---

<sup>27</sup> Profiles in an elimination database are stored separately and cannot be loaded onto or searched against the National DNA Database.

the enquiry. Such events should automatically trigger further investigation, so that any lessons learnt can be used to improve the resilience of anti-contamination procedures in the future.

- (v) Consumables, such as plastic tubes and cotton swabs, should be chemically pre-treated and certified DNA free, not merely 'sterile' (i.e. uncontaminated by bacteria). Batches of consumables should be monitored for the presence of contaminant DNA.

Failure to implement these essential anti-contamination measures may have very serious consequences for the administration of criminal justice. One particularly memorable cautionary tale comes from Germany. Several years ago, DNA from the same female donor was detected in some 30 stains from multiple crime scenes, including the murder of a policewoman in Heilbronn. The media christened this mysterious and improbably prolific offender 'The Phantom of Heilbronn' (BBC News Channel, 2008 and 2009). But the only real phantom in this story was contamination. The German police later discovered that the Phantom's DNA was present in cotton swabs used to collect DNA samples, and the origin of the DNA was eventually traced back to a woman working in a cotton swab factory.

## 6.8 *Issues of interpretation*

LTDNA profiles are more difficult to interpret than standard profiles. They tend to suffer from a higher incidence of the profiling artefacts discussed in previous sections (Whitaker et al., 2001), including:

- ***heterozygote imbalance***. LTDNA profiles may display greater imbalance in a pair of peaks from a heterozygote donor than in standard profiling.
- ***allelic dropout***. LTDNA profiles have a higher than normal incidence of 'missing' alleles, i.e. incomplete profiles, a phenomenon known as 'allelic dropout'.
- ***additional alleles; 'drop-in'***. Conversely, an LTDNA profile sometimes contains smaller peaks in addition to the peaks of the donors. These additional 'drop-in' alleles derive from very small quantities of contaminant DNA, from



the crime scene or elsewhere, which has been amplified and detected in the profiling process. Occasionally, peaks are detected from unused plastic tubes supposedly containing only the multiplex chemicals. One explanation for this ‘drop in’ phenomenon is that tubes have become contaminated by DNA floating in the air of the laboratory. The anti-contamination procedures described above are designed to minimise drop-in.

Drop-in must be distinguished from *gross contamination*, where large amounts of DNA are deposited at some stage in the process of producing a profile and additional peaks at multiple loci or a full DNA profile are detected. For example, DNA transferred to tubes during their manufacture would be gross contamination. Elimination databases address gross contamination, but single additional alleles cannot be systematically monitored or traced to their source using this method.

- **Stutters.** Comparatively large stutters are routinely observed in the EPG of an LTDNA profile. For example, it is possible to observe a *stutter ratio* of 25% in 34-cycle SGMPlus profiles.

The susceptibility of LTDNA profiles to artefacts of various kinds underscores the significance of the skill and experience of individual forensic scientists in interpreting profiling results and assessing their potential probative value. The Court of Appeal has rejected the argument that partial profiles must necessarily be excluded from a criminal trial, stating that ‘the fact that there exists in the case of all partial profile evidence the possibility that a ‘missing’ allele might exculpate the accused altogether does not provide sufficient grounds for rejecting such evidence’.<sup>28</sup>

6.9 Replication of analysis is employed to assist in the assessment and evaluation of LTDNA profiles produced by LCN LTDNA systems. For example, each sample may be divided into three portions and amplified twice, with the third sample being retained for later use should any analytical problem arise with the first two

---

<sup>28</sup> *R v Bates (Richard)* [2006] EWCA Crim 1395, [30].

amplifications or if there is any ambiguity in the results. The third sample can also be used for re-analysis using other tests or by other forensic providers or defence experts.

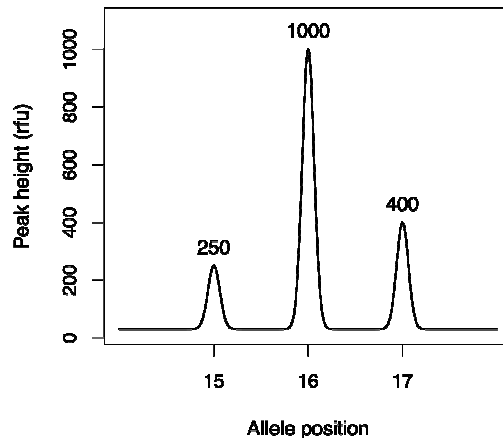
#### 6.10 *Assessing the probative value of LTDNA profiles*

Previous sections of this *Guide* have described and illustrated a logical method for assessing the quantified probative value of DNA profiles, utilising likelihood ratios and paired propositions. We first applied this method to questioned profiles deriving from a single donor. This enabled us to introduce standard features of the DNA profiling process including allele databases, ethnic appearance populations, co-ancestry allowance and sampling allowance. We then applied the same logical approach to two person questioned profiles, introducing the additional concepts of allele masking and mixing proportion. Each of these illustrations was predicated on the assumption that there was sufficient DNA to generate a full profile or profiles.

We now illustrate the calculation of likelihood ratios for profiles generated from very small amounts of DNA using the LTDNA process. The following three examples show how forensic scientists address interpretational issues that commonly arise in practice, including unbalanced peaks and large stutters, allelic dropout and additional alleles. For ease of exposition, we will consider each aspect sequentially, though in real case-work a forensic scientist might have to grapple with two or more of these interpretational issues concurrently. Methods for assigning weight to LTDNA profiles will also be discussed.

#### 6.11 *Example #1: One-person profile with larger heterozygote balance and stutter ratio*

Figure 6.2 depicts one locus (D3) of a profile in a case where the defendant genotype is known to be 16,17 at D3; and the questioned profile, produced using SGMPlus at 34 amplification cycles, consists of three peaks in positions 15, 16 and 17 with corresponding heights 250 rfu, 1000 rfu and 400 rfu, respectively



**Figure 6.2 LTDNA questioned profile at locus D3, produced with SGMPlus at 34 cycles**

Suppose that the forensic scientist has examined the questioned profile across all loci, before looking at the defendant's genotype, and concluded that it is a single person profile. (In reality, this assessment would require consideration of many other factors, potentially including the implications of alternative scenarios advanced by the defence postulating more than one donor, but we ignore these complexities here: our focus is on the analytical method.)

The forensic scientist then compares the questioned profile against the defendant's genotype and is satisfied that the defendant is a possible donor. To arrive at this judgement, the forensic scientist considers the balance between peaks 16 and 17 and the size of the stutter (peak 15) with respect to its parent peak. Although the peaks seem unbalanced with a high stutter ratio, LTDNA systems are known to produce profiles of this kind. Replication of analysis might be useful in this scenario; but we set these operational issues to one side for present purposes.

6.12 The competing propositions to be considered are, as before:

***Prosecution Proposition*** (PP): the DNA recovered from the crime scene came from the accused.

***Defence Proposition*** (DP): the DNA recovered from the crime scene came from an unknown person, not blood related to the accused.

The likelihood ratio is the ratio of the two probabilities, of obtaining the evidence, E, if the prosecution's proposition is true,  $p(E \mid PP)$ ; and of obtaining the evidence, E, if the defence proposition is true,  $p(E \mid DP)$ . The numerator of the likelihood ratio is given by  $p(E \mid PP)$ ; whilst  $p(E \mid DP)$  supplies the denominator.

In this example the evidence, E, is the conjunction of two 'events': the crime profile ( $E_1$ ) and the accused's reference profile ( $E_2$ ). Applying the laws of probability (see n.22, para.4.12, above), the likelihood ratio becomes the ratio of  $p(E_1 \mid E_2, PP)$  to  $p(E_1 \mid E_2, DP)$ .

6.13 The defence proposition nominates another person as the donor. Assuming a single donor, the only eligible genotype is 16,17.

The likelihood ratio  $p(E_1 \mid E_2, PP) / p(E_1 \mid E_2, DP)$  is then calculated in the following way:

$$\text{Likelihood ratio} = \frac{\text{p1: probability of obtaining the questioned profile given that the donor is 16,17}}{\text{p2: probability of obtaining the questioned profile given that the donor is 16,17}} \times \text{p3: the probability of finding an unknown person at random whose genotype is 16,17 in a population of people not blood related to the accused}$$

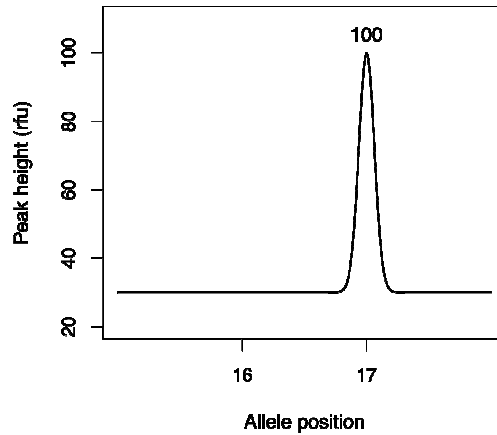
Notice that both the numerator and denominator of the likelihood ratio contain the probability of obtaining the questioned profile, on the assumption that the donor's genotype is 16,17. Probabilities p1 and p2 are thus identical and cancel out in the equation, simplifying the likelihood ratio to:

$$\text{Likelihood ratio} = \frac{1}{\text{p3: the probability of finding an unknown person at random whose genotype is 16,17 in a population of people not blood related to the defendant}}$$

Probability p3 is a genotype probability, and can be calculated in the usual manner, elucidated in Part 4 of this *Guide*.

6.14 *Example #2: One-person profile with possible allelic dropout*

Figure 6.3 depicts a questioned profile, produced using SGMPlus at 28 cycles, consisting of just a single peak 17 in locus D3 with a peak height of 100 rfu.



**Figure 6.3 LTDNA questioned profile at locus D3, produced with SGMPlus at 28 cycles**

LTDNA profiles can be produced at 28 instead of 34 amplification cycles, on current protocols. The quantification stage of the profiling process, described in Part 2, may provide information useful to a forensic scientist in interpreting an LTDNA profile.

Our background assumptions are the same as in the previous example. The defendant's genotype in locus D3 is known to be 16,17. The forensic scientist has examined the questioned profile across all loci and is satisfied that it came from a single donor. The forensic scientist then satisfies herself that the defendant is a possible donor of the profile. It would be unusual to obtain only one peak from a heterozygote donor in a profile produced from an optimal amount of DNA, but this is routinely encountered in a LTDNA profile.

6.15 With these assumptions in place, the probabilistic value of the profile may be quantified. Our now-familiar competing propositions are:

**Prosecution Proposition (PP):** the DNA recovered from the crime scene came from the accused.

**Defence Proposition (DP):** the DNA recovered from the crime scene came from an unknown person, not blood related to the accused.

The likelihood ratio is the ratio of the two probabilities, of obtaining the evidence, E, if the prosecution's proposition is true,  $p(E | PP)$ ; and of obtaining the evidence, E, if the defence proposition is true,  $p(E | DP)$ . The numerator of the likelihood ratio is given by  $p(E | PP)$ ; whilst  $p(E | DP)$  supplies the denominator.

The evidence, E, consists of the crime profile,  $E_1$ , and the defendant's reference profile,  $E_2$ . Using the same probabilistic derivation that we have been employing all along, the likelihood ratio becomes  $p(E_1 | E_2, PP) / p(E_1 | E_2, DP)$ .

6.16 The numerator probability is the probability of obtaining the questioned profile given that the donor genotype is 16,17.

The defence proposition nominates an unknown person not blood related to the accused as the donor of the profile. Consistent with the profiling result, the unknown donor could have a range of genotypes, such as 17,17; or 16,17; or any genotype 17,q, where q represents an allele in locus D3 other than 16 or 17.

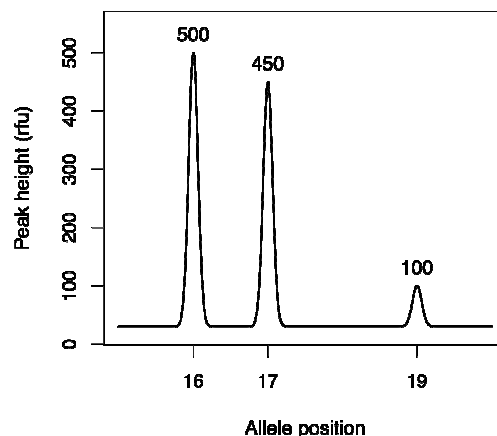
The likelihood ratio is therefore:

$$\begin{aligned}
 & \text{p1: probability of obtaining the questioned profile given that the} \\
 & \qquad \qquad \qquad \text{donor is 16,17} \\
 & \hline
 & \text{p2: probability of obtaining the questioned profile given that the} \\
 & \qquad \qquad \qquad \text{donor is 17,17} \\
 & \qquad \qquad \qquad \times \\
 & \text{p3: the probability of finding an unknown person at random} \\
 & \qquad \text{whose genotype is 17,17 in a population of people not blood} \\
 & \qquad \text{related to the accused} \\
 & \qquad \qquad \qquad + \\
 \text{Likelihood ratio} = & \text{p4: probability of obtaining the questioned profile given that the} \\
 & \qquad \qquad \qquad \text{donor is 16,17} \\
 & \qquad \qquad \qquad \times \\
 & \text{p5: the probability of finding an unknown person at random} \\
 & \qquad \text{whose genotype is 16,17 in a population of people not blood} \\
 & \qquad \text{related to the accused} \\
 & \qquad \qquad \qquad + \\
 & \text{p6: probability of obtaining the questioned profile given that the} \\
 & \qquad \qquad \qquad \text{donor is 17,q} \\
 & \qquad \qquad \qquad \times \\
 & \text{p7: the probability of finding an unknown person at random} \\
 & \qquad \text{whose genotype is 17,q, in a population of people not blood} \\
 & \qquad \text{related to the accused}
 \end{aligned}$$

6.17 In contrast to the previous examples we have considered, the probability of obtaining the questioned profile if the donor's genotype is 16,17, which still appears in both the numerator and the denominator of the likelihood ratio, does not cancel through. The calculation cannot be simplified to produce a numerator of 1, as we did before. In this situation the likelihood ratio can only be calculated by assigning values to each of the relevant probabilities,  $p_1$  to  $p_7$ .

6.18 ***Example #3: One-person profile with an additional small peak***

For our third example, consider a case where the defendant's genotype at locus D3 is still 16,17, but this time the questioned profile (produced using SGMPlus run at 34 amplification cycles) consists of peaks 16, 17 and 19 at locus D3 with corresponding peak heights 500 rfu, 459 rfu and 100 rfu, respectively. The relevant portion of the EPG is shown in Figure 6.4.



**Figure 6.4 Questioned profile at locus D3 with an additional peak**

Additional alleles of this type are known to occur for LTDNA single person profiles (Whitaker et al., 2001).

Once again, we make the simplifying assumption (ignoring possibilities of allelic dropout and multiple donors etc., that would be confronted in real case-work) that the forensic scientist has examined the profile across all loci and is satisfied that the profile derives from a single donor.

6.19 In order to calculate a likelihood ratio for this profile, the forensic scientist would need to assign a value to the probability of obtaining the questioned profile at locus D3 given that the donor's genotype is 16,17. This could only be done by drawing on relevant scientific literature and the scientist's own personal experience of interpreting DNA profiling results. We have reached the outer limits of LTDNA theory and practice as currently understood. The forensic scientist must disclose her results, even if they are not, in her opinion, amenable to further evaluation.

6.20 *Assessing the probative value of LTDNA profile evidence*

It is evident from the foregoing discussion that interpreting LTDNA profiles requires skill, professional experience, and scientific judgement at various stages of the analytical work. Likelihood ratios can be calculated only on the basis of various assumptions that require independent substantiation. Many questions at the extremities of LTDNA technology remain unanswered, and scientific disputes between experts are sometimes ventilated in litigation.<sup>29</sup> The Forensic Regulator is currently investigating the feasibility of providing further guidance.

Research investigating drop-out and drop-in probabilities is on-going (Gill et al., 2000). This has stimulated further work on estimates of dropout probabilities from profile data (Tvedebrink et al., 2009, Tvedebrink et al., 2012) and new approaches to coping with stutters (Balding and Buckleton, 2009). Other researchers have modelled peak heights, heterozygote balance, allelic dropout and stutters employing user-defined settings derived from scientific experience (Cowell et al., 2011) or settings estimated from quantitative data (Puch-Solis et al., 2012).

6.21 In *R v Reed and Reed* the Court of Appeal indicated, as a general rule of thumb, that LTDNA profiles derived from DNA weighing 200 picograms or more would be admissible. The Court did not categorically rule out the admissibility of profiles obtained from 100-200 picograms, but warned:<sup>30</sup>

---

<sup>29</sup> See, e.g., *R v Hookway and Noakes* [2011] EWCA Crim 1989.

<sup>30</sup> *R v Reed and Reed*; *R v Garmson* [2010] 1 Cr App R 23, [74].



There may be cases where reliance is placed on a profile obtained where the quantity of DNA analysed is within the range of 100–200 picograms where there is disagreement on the stochastic threshold on the present state of the science. We would anticipate that such cases would be rare and that, in any event, the scientific disagreement will be resolved as the science of DNA profiling develops. If such a case arises, expert evidence must be given as to whether in the particular case, a reliable interpretation can be made. We would anticipate that such evidence would be given by persons who are expert in the science of DNA and supported by the latest research on the subject.

Subsequently, in *R v Thomas*,<sup>31</sup> an expert instructed by the defence initially concluded that there was too little DNA in the relevant sample to produce meaningful results. Following a pre-trial meeting with the Forensic Science Service (FSS) scientist in the case, within the framework of the Criminal Procedure Rules, both experts agreed that there was just enough DNA to produce an LTDNA profile, but insufficient for the purposes of quantification. In other words, the amount of DNA available fell below the '*stochastic threshold*'. The Court of Appeal expressed some reservations about the FSS forensic scientist's use of unpublished test results,<sup>32</sup> but confirmed that the profile itself was admissible. The forensic scientist was entitled to say, on the basis of her substantial forensic experience, that the LTDNA profile supported the prosecution's case (specifically, the proposition that DNA recovered from a gun could have come from the accused), without attempting – or indeed being able – to quantify the degree of support it provided.

6.22 On the particular facts of *Thomas*, where the DNA evidence was ultimately regarded as a mere 'sideshow' in a compelling circumstantial case against the accused, the Court of Appeal rejected the defence's contention that the FSS scientist should have been restricted to saying that the LTDNA profile 'could not exclude' the accused as a potential donor. In view of the course of testimony at trial, the Court of Appeal concluded, these alternative formulations amounted to 'a distinction without a

---

<sup>31</sup> *R v Thomas* [2011] EWCA Crim 1295.

<sup>32</sup> 'The difficulty about the simulation experiments in this case is not that they were unpublished but that [the FSS expert] seems to have known virtually nothing about them beyond the bare statement in the FSS manual that "Unpublished simulation experiments have shown that it is rare to observe all twenty alleles by chance". Taken by itself, that would provide an extremely thin basis for [the expert's] statement of opinion about the significance of the DNA results': *ibid.* [38].

difference'. This brings us to the vital question of how the results of DNA profiling should be communicated to jurors in criminal trials.

## 7. Presenting DNA Evidence in the Courtroom

7.1 The previous Parts of this *Guide* have set out a logical approach to quantifying the probative value of DNA (including LTDNA) evidence. The general approach, which we first introduced in *Practitioner Guide No 1*, involves calculating likelihood ratios for the probability of evidence conditioned on pairs of mutually exclusive propositions, e.g. the probability of obtaining a given DNA profile if the accused is the source of the DNA ('the prosecution proposition'), divided by the probability of obtaining that profile if some other person not blood related to the accused is the source ('the defence proposition'). It is not necessary for the defence actually to advance any affirmative proposition in order to calculate these likelihood ratios. The 'defence proposition' is simply the negation of the 'prosecution proposition', which in turn is a function of the relevance of the evidence in purporting to prove the accused's guilt.

Likelihood ratios are a strictly rational and mathematically validated mechanism for quantifying evidential weight or probative value, i.e. *the strength* of evidential support for a particular proposition. They are employed by many forensic scientists in their case-work. Although likelihood ratios also feature in Bayes' Theorem, there is nothing inherently or distinctively 'Bayesian' about the use of likelihood ratios or the importance of considering the probability of evidence under competing propositions. It is simply a matter of elementary logic that evidence compatible with guilt could also be compatible with innocence, and one cannot therefore assess its relevance or probative value without first considering how a particular item of evidence might bear on *both* sides of the argument, for and against. This inquiry is inescapably probabilistic.

7.2 It does not necessarily follow that the analytical results of forensic testing should be presented in criminal trials in their 'raw' form. Part of the role of expert witnesses is to mediate between scientific findings and lay understandings. Forensic scientists preparing court reports or testifying in the courtroom as expert witness should strive to make their findings accessible to judges and lay jurors and to assist the court to form an appropriate assessment of the probative value of scientific evidence. Forensic scientists' assessments of probative value are always partial and provisional, pending

the jury's ultimate, holistic determination, but the forensic scientist presenting DNA evidence must first arrive at her own conclusions - utilising the techniques described in this *Guide* – before being in a position to try to communicate the meaning and value of her evidence to the jury.

Presentation is pivotal. Common sense tells us that the way in which evidence is presented to the fact-finder might be more or less conducive to its appropriate evaluation. Some forms of presentation may be relatively clear and informative, whilst others might be especially prone to misinterpretation or to confusing or misleading the fact-finder. Some examples of misleading phraseology, such as 'is consistent with', were discussed in *Practitioner Guide No 1*. The comparative merits of 'provides support' as against 'could not exclude' were debated in the *Thomas*<sup>33</sup> case, mentioned in the previous Part.

7.3 DNA profiling evidence can be presented in different ways, with or without numbers, and in alternative probability formats (Redmayne, 2001: ch. 4), posing the question as to which mode of presentation is best calibrated to make the relevance and probative value of the evidence transparent to jurors. Lay people may not grasp the finer points of probability theory, or even the basics.

Common sense expectations are supported by behavioural science data. It has been shown, for example, that mock jurors are inclined to assess the probative value of evidence differently when it is expressed in mathematically identical terms (Koehler, 2001). This is irrational, from a logical point of view, and implies limited numeracy in the general population. As Lindsey et al. (2003: 154) put it, '[e]xpressions of probabilities that are mathematically equivalent are not necessarily psychologically equivalent'.

Interestingly, some studies have concluded that lay people find likelihood ratios especially difficult to interpret and prefer to deal with discrete probabilities or natural frequencies (see Taroni and Aitken, 1998a and 1998b). Koehler (1996: 877) suggests that '[e]ven when likelihood ratios are properly conveyed, there is little reason to believe that jurors will understand what they mean and how they should be used.

---

<sup>33</sup> *R v Thomas* [2011] EWCA Crim 1295.

Although they have scientific merit, likelihood ratios... are not easy to understand'. This lends some support to the position of the English Court of Appeal in stipulating the form in which DNA evidence should be presented in court.

#### 7.4 *The Ruling in Doheny and Adams*

In *R v Doheny and Adams*<sup>34</sup> the Court of Appeal stipulated that expert witnesses giving evidence of DNA profiling results should confine their testimony to what the Court described as 'the random occurrence ratio', but which is more commonly known as the *random match probability* (RMP). Expert testimony framed in terms of the random occurrence ratio/RMP would enable the trial judge to direct the jury along the following lines:

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 [crime] stain 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 that other small group of men who share the same DNA characteristics.<sup>35</sup>

Under the *Doheny and Adams* approach, random match probabilities displace likelihood ratios in courtroom testimony. The 'four or five white males' translation reflects the (comparatively limited) discriminating power of DNA profiling technology at the time. Today, the standard RMP for a full profile is 1-in-a-billion. In fact, it is apparent from other Court of Appeal judgments, as well as from anecdotal experience, that some forensic scientists have continued to present their evidence in terms of likelihood ratios, *Doheny* notwithstanding.<sup>36</sup>

7.5 If members of the public are, generally speaking, more familiar and comfortable with probabilities than with numerical likelihood ratios, it is reasonable to suppose that jurors will be better equipped to assess the probative value of DNA profiling evidence expressed as an RMP; especially if trial judges further spell out the logical

---

<sup>34</sup> *R v Doheny and Adams* [1997] 1 Cr App R 369, CA.

<sup>35</sup> *ibid.* 375.

<sup>36</sup> See, e.g., *R v Britten* [2008] EWCA Crim 312; *R v C (Danny)* [2002] EWCA Crim 1979, [5]. And in Scotland: *Hopton v HM Advocate* [2010] HCJAC 10, 2010 SCL 652, [12].

implications of the RMP, as the Court of Appeal encouraged them to do in *Doheny and Adams*.

However, this still assumes that jurors can make sense of the RMP, as a quantified measure of probative value, and relate it to the other evidence in the case. Given that a full SGMPlus profile now produces an RMP of the order of one-in-a-billion, what would jurors make of the notion that, *if the DNA evidence is considered in isolation*, there are perhaps another seven people currently alive in the world today who might share the same profile as the accused?

7.6 Enduring doubts about the ability of lay jurors to make much sense of RMPs (or other quantified measures of probative value) have prompted some forensic scientists to develop numerical scales to translate calculated likelihood ratios into verbal descriptions of evidential strength, on the assumption that such verbal descriptions will better convey the true probative value of evidence to lay fact-finders. Here is one example of such a scale, which has been employed in casework by some FSS scientists and is endorsed by the Association of Forensic Science Providers (AFSP) (2009):

<b>Value of likelihood ratio</b>	<b>Verbal equivalent</b>
>1-10	Weak or limited support
10-100	Moderate support
100-1000	Moderately strong support
1000-10,000	Strong support
10,000-1,000,000	Very strong support
>1,000,000	Extremely strong support

According to this scale of verbal equivalents, a full DNA profile would constitute ‘extremely strong support’ for the prosecution proposition. (In fact, it exceeds this threshold by three orders of magnitude, but it is hard to see how one can top ‘extremely strong support’ as a verbal expression of probative value.)

7.7 In a recent case concerned with footwear marks, the Court of Appeal expressed strong reservations about the practice of translating likelihood ratios into verbal descriptors of evidential strength.<sup>37</sup> The logic of the process does not appear to have been exhaustively canvassed in the proceedings, and the better view may be that this decision is confined to its facts (see Redmayne et al., 2011). Verbal descriptors of evidential strength have appeared in subsequent criminal appeals without attracting the Court's censure.<sup>38</sup>

In relation to non-quantified evidence, there is no alternative to verbal descriptions of one kind or another. The Court of Appeal appears to be entirely comfortable with the notion that an expert in fields requiring significant elements of subjective interpretation, such as facial-mapping<sup>39</sup> and handwriting analysis,<sup>40</sup> can quantify the strength of their evidence in terms of a verbal scale rooted in the expert's own professional judgement and experience.<sup>41</sup>

A similar issue arises in relation to LTDNA profiles below the stochastic threshold. The Court of Appeal in *R v Thomas* thought it 'troubling' that the forensic scientist could only say that an LTDNA profile 'provide[d] support' for the view that the appellant was a contributor to the DNA recovered from the pistol' and that 'it was 'rare' to find all 20 components of a DNA profile by chance'.<sup>42</sup> These expressions were not contextualised in terms of any sliding scale of evidential strength, and this was one reason why the expert's opinion 'was so unsatisfactory'.<sup>43</sup> Nonetheless, as a

---

<sup>37</sup> *R v T* [2011] 1 Cr App R 9, [2010] EWCA Crim 2439.

<sup>38</sup> See, e.g., *R v South* [2011] EWCA Crim 754 (where a forensic scientist testified that footwear mark evidence provided 'moderately strong support' for the proposition that D's shoe was the source of the mark, apparently relying on virtually identical factors to those considered by the expert in calculating likelihood ratios in *R v T*).

<sup>39</sup> *R v Atkins* [2010] 1 Cr App R 8, [2009] EWCA Crim 1876; *R v Bull* [2010] EWCA Crim 2542.

<sup>40</sup> *R v Bilal* [2005] EWCA Crim 1555.

<sup>41</sup> And see *R v Shillibier* [2006] EWCA Crim 793, [2007] Crim LR 639, where the expert in soil analysis also had some statistical basis for his opinion.

<sup>42</sup> *R v Thomas* [2011] EWCA Crim 1295, [35].

<sup>43</sup> *Ibid.* [37].

general proposition ‘an expert assessment based on experience may be admissible even in the absence of a statistical evaluation of likelihood, provided that the matter is approached with suitable caution and the nature of the assessment is made crystal clear to the jury’.<sup>44</sup> These criteria were found to have been satisfied on the facts of *Thomas*, where the DNA evidence was anyway regarded as incidental to the accused’s conviction.

7.8 Exploring the broader ramifications of verbal expressions of evidential weight for scientific evidence in general would take us beyond the scope of our present focus on DNA evidence. The ruling in *Doheny and Adams*, that DNA profiling evidence should be expressed (only) in terms of RMPs, has never subsequently been questioned or doubted by an English court, though the extent of compliance by forensic scientists in first instance trials is impossible to gauge.

It makes sense to talk about ‘random match probabilities’ in relation to single person profiles produced by standard STR profiling techniques. Those lawyers who regularly encounter DNA evidence in practice have presumably become somewhat familiar with this way of understanding the probative value of DNA profiles, and it does not appear to have been a topic of protracted legal argument in the decade since *Doheny and Adams* was decided.

Unfortunately, the concept of a RMP is not applicable to mixed profiles or LTDNA profiles. The more accurate and flexible concept is ‘genotype probability’. Multiple genotype probabilities can be assigned and factored into the calculation of likelihood ratios for mixed profiles and in LTDNA profiling, utilising the formulae and probabilistic methods described in this *Guide*. The RMP also has limited application to activity-level propositions, as *Practitioner Guide No 4* will elucidate.

Courts in England and Wales, Scotland and Northern Ireland have yet to confront the limited applicability of random match probabilities, presumably because it does not arise in relation to single person STR profiles. However, if mixed and LTDNA profiles are increasingly received in evidence in criminal proceedings (as appears

---

<sup>44</sup> Ibid.



likely), it may be necessary to revisit the question of likelihood ratios, and their expression through verbal equivalents, in relation to DNA evidence.

### 7.9 *Illegitimately transposing the conditional ('the prosecutor's fallacy')*

Owing to its explicit quantification, DNA profiling evidence has been especially susceptible to the probabilistic reasoning error popularly known as 'the prosecutor's fallacy', but more technically and accurately described as illegitimately *transposing the conditional*. *Practitioner Guide No 1* introduced this as a prevalent 'trap for the unwary', and explained how the conditional could be transposed *legitimately* utilising Bayes' Theorem.

Any participant in criminal proceedings –lawyers, judges, jurors, or forensic scientists – might fall into the trap of illegitimately transposing the conditional. Prosecutors are by no means the only culprits. 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.

7.10 The most direct way of conceptualising the error is to say that it confuses ('transposes') the event on which particular probabilities are conditioned. 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!<sup>45</sup> Proposition #1 is true, whereas proposition #2 is false. Moreover, proposition #2 patently does not follow from proposition #1. When criminal justice professionals illegitimately

---

<sup>45</sup> 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*'.

transpose the conditional they perpetrate an error equivalent to treating proposition #1 as though it were the same as or implied by proposition #2.

7.11 In the context of criminal proceedings, the gross form of the error confuses the probability of the evidence assuming innocence,  $p(E | I)$ , with the probability of innocence assuming the evidence,  $p(I | E)$ . Yet, as we just saw from the monkey example, these are patently not equivalent quantities. Put another way:

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

Mathematical notation is particularly useful here, because we can see that ‘E’ and ‘I’ have migrated and 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.

7.12 Throughout this *Guide*, we have been at pains to differentiate between the prosecution proposition (e.g., the DNA in the questioned profile came from the accused) and the likelihood ratio, which is a measure of the (potential) probative value of the evidence in the light of two competing propositions. The first stage in calculating a likelihood ratio for DNA profiling evidence is to formulate a pair of propositions relevant to the disputed issue(s) in the case. Profiling evidence can then be assessed, first as a question of analytical interpretation (taking account of the potential number of donors, the amount of DNA available for testing, possible artefacts, etc.), and then by calculating the ratio of the probabilities of the evidence under each proposition in the pair.

It is when forensic scientists (or lawyers or judges) depart from this strictly logical analysis – perhaps in the cut-and-thrust of cross-examination – that illegitimate transpositions of the conditional are liable to occur.

7.13 This is how the fallacy was perpetrated in *R v Deen*,<sup>46</sup> an early DNA case where the profile was calculated to have an RMP of 1 in 3 million:

*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.

The fundamental mistake is contained in counsel's first question, which asks the expert about the probability of D's being the source of the profile, which is the prosecution's proposition –  $p(PP | E)$ , rather than the random match probability; which is the probability of the evidence assuming the defence proposition, that D is *not* the donor:  $p(E | DP)$ . Alternatively, this could have been correctly expressed as the likelihood ratio of the evidence:  $p(E | PP) / p(E | DP)$ . In either reformulation the expert would be testifying to *the probability of the evidence*, not to the probability of any proposition, whether the prosecution's or the defence's. It is the jury's job, not the expert's, to decide whether or not particular factual propositions have been established by the evidence.

Having been asked the wrong question, the expert in *Deen* then confounded the fallacy, even to the extent of pronouncing himself 'sure' that D was the source of the stain. In fact, an RMP of 1 in 3 million implies that about 20 people in the UK would be expected to share the same profile.

---

<sup>46</sup> *R v Deen*, CA, *The Times*, 10 January 1994.

7.14 An expert witness called by the prosecution likewise illegitimately transposed the conditional in *Doheny and Adams*, as recounted by the Court of Appeal:<sup>47</sup>

‘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.’

Again, the rot starts with counsel’s first question. An RMP of 1 in 40 million means that the probability of selecting a person at random with a matching profile is 1 in 40 million, i.e.  $p(E | DP)$ . This is patently not the same value as  $p(DP | E)$ , the probability that the defence proposition is true, and the accused is therefore not the source of the DNA in the questioned profile, in light of the evidence. The conditional has been illegitimately transposed.

7.15 These ‘classic’ illustrations of illegitimate transposition date from the relatively early days of DNA profiling. But there is every reason to think that this pervasive problem endures – partly owing to an embarrassment of anecdotal examples (such as those mentioned by Cooke, 2007), but also because thinking in the right way, about the probability of evidence rather than the probability of propositions, does not seem to come naturally to most people and it is very easy to slip into error. Thus, in a recent case *Laws LJ* stated, rehearsing material facts: ‘An incomplete DNA profile found on a swab taken from the magazine of the gun matched the appellant’s DNA. The chance of it being from someone else was one in a million’.<sup>48</sup> This is a blatant illegitimate transposition of the conditional. If the random match probability was correctly calculated as 1-in-a-million, this means that one person in every million in the suspect population would be expected to match the questioned profile. Assuming for the sake of argument that there are eight million people in the greater London area and that this is the relevant suspect population, one would expect there to be eight people, plus the accused, with a matching profile. In other words, the probability of somebody other

---

<sup>47</sup>*R v Doheny and Adams* [1997] 1 Cr App R 369, 377-8, CA.

<sup>48</sup>*R v Cambridge (Conor)* [2011] EWCA Crim 2009, [2].

than the accused being the source of the profile, taking only the profiling evidence into account, is  $8/9 = 0.89$  or 89% (rounded up to two decimal places).

- 7.16 The Court of Appeal in *Doheny and Adams* had the ‘prosecutor’s fallacy’ firmly in mind when it admonished forensic scientists testifying about DNA profiling evidence to confine themselves to stating the RMP. The Court of Appeal insisted that a scientist ‘should not be asked his opinion on the likelihood that it was the defendant who left the crime stain, nor when giving evidence should he use terminology which may lead the jury to believe that he is expressing such an opinion’.<sup>49</sup>

If experts follow this injunction, they are less likely to stumble into illegitimate transpositions of the conditional. Limiting themselves to expressing the RMP is a viable strategy for single person STR profiles but will not work, as we have seen, in relation to mixed profiles or LTDNA profiles. This serves only to reinforce the importance of cultivating and internalising a strictly logical approach to assessing the probative value of profiling results. If lawyers and courts are able to grasp the fundamentals of the calculations presented in this *Guide*, they too can play an active role in policing, or better still pre-empting, illegitimate transpositions of the conditional.

- 7.17 Even if experts themselves are fastidious in the language of their report writing and testimony it does not necessarily follow, of course, that lawyers, or trial judges, or juries will avoid the error, without further careful guidance. It would in theory be possible to teach jurors to calculate likelihood ratios in the same way that many forensic scientists currently do. This would, of course, be a major departure from traditional trial practice, and the Court of Appeal strongly deprecated any developments in this direction in *R v Adams*,<sup>50</sup> where the defence had attempted to instruct the jury in the use of Bayes’ Theorem:

[W]e regard the reliance on evidence of this kind... as a recipe for confusion, misunderstanding and misjudgment, possibly even among counsel, but very probably among judges and, as we conclude, almost certainly among jurors. It would seem to us that this was a case properly approached by the jury

---

<sup>49</sup> *R v Doheny and Adams* [1997] 1 Cr App R 369, 374.

<sup>50</sup> *R v Adams (No.2)* [1998] 1 Cr App R 377, CA.

along conventional lines.... We do not consider that [juries] will be assisted in their task by reference to a very complex approach which they are unlikely to understand fully and even more unlikely to apply accurately, which we judge to be likely to confuse them and distract them from their consideration of the real questions on which they should seek to reach a unanimous conclusion. We are very clearly of opinion that in cases such as this, lacking special features absent here, expert evidence should not be admitted to induce juries to attach mathematical values to probabilities arising from non-scientific evidence adduced at the trial.<sup>51</sup>

Those of a Bayesian disposition might be tempted to interpret these remarks as a victory for the dark forces of ignorance over the light of science. This would be hasty and excessively pessimistic conclusion. Juries are empanelled in order to inject common sense reasoning into criminal adjudication. But this does not mean that criminal trials are a forensic free-for-all. Both the content of the information presented to juries, and the manner of its presentation, are carefully regulated by the law of criminal evidence and procedure. To this extent, the jury's common sense reasoning is constrained and channelled in conformity with the rule of law. The question, then, is whether anything further can be done, within the framework of traditional criminal trial procedure, to facilitate the jury's informed evaluation of DNA evidence?

7.18 Assuming that DNA evidence has been presented accurately and clearly by forensic scientists in their evidence in-chief, the onus is then on trial counsel and, especially, the judge to maintain these standards of clarity for the benefit of the jury. Needless to add, lawyers and judges will better succeed in their respective roles if they themselves have a firm grasp of the probabilistic foundations of DNA evidence and can avoid fallacious reasoning.

Judicial summings-up should be, above all, even-handed and helpful to the jury. This involves contextual judgments, tailored to the facts of the case. DNA evidence will not be excluded from the trial just because there is a bona fide disagreement between experts on a point of interpretation.<sup>52</sup> Nor is there a special form of judicial direction that must be given whenever experts disagree.<sup>53</sup> The Court of Appeal consistently emphasises the importance of viewing DNA evidence within the context of the case as

---

<sup>51</sup> *Ibid.* 385.

<sup>52</sup> *R v Hookway and Noakes* [2011] EWCA Crim 1989.

<sup>53</sup> *R v Thomas* [2011] EWCA Crim 1295, [44].

a whole, and this is what trial judges should encourage jurors to do when assessing the probative value of DNA profiling evidence in the light of the criminal burden and standard of proof.

## Appendix A – Bibliography

- Aitken, C., Roberts, P. and Jackson, G. (2010) *Fundamentals of Probability and Statistical Evidence in Criminal Proceedings*. London: RSS, 2010), [www.rss.org.uk/statsandlaw](http://www.rss.org.uk/statsandlaw).
- 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, West Sussex: Wiley.
- Balding, D.J. and Buckleton, J. (2009) ‘Interpreting Low Template DNA Profiles’ 4 *Forensic Science International: Genetics* 1.
- Balding, D.J. and Nichols, R.A. (1994) ‘DNA Profile Match Probability Calculation: How to Allow for Population Stratification, Relatedness, Database Selection and Single Bands’ 64 *Forensic Science International* 125.
- Bramley, B. (2009) ‘DNA Databases’ in J. Fraser and R. Williams (eds), *Handbook of Forensic Science*. Cullompton, Devon: Willan Publishing.
- Buckleton, J., Triggs, C.M. and Walsh, S.J. (eds) (2005) *Forensic DNA Evidence Interpretation*. Boca Raton, Florida: CRC Press.
- Caddy, B., Taylor, G.R., and Linacre, A.M.T. (2008) *A Review of the Science of Low Template DNA Analysis*. London: Forensic Science Regulator. [www.homeoffice.gov.uk/publications/police/operational-policing/](http://www.homeoffice.gov.uk/publications/police/operational-policing/)
- Clayton, T., Whitaker, J., Sparkes, R., and Gill, P. (1998) ‘Analysis and Interpretation of Mixed Forensic Stains Using DNA STR Profiling’ 91 *Forensic Science International* 70.
- Cole, S.A. (2005) ‘More Than Zero: Accounting for Error in Latent Fingerprint Identification’ 95 *Journal of Criminal Law and Criminology* 985.
- Cole, S.A., and Dioso-Villa, R. (2009) ‘Investigating the “CSI Effect” Effect: Media and Litigation Crisis in Criminal Law’ 61 *Stanford Law Review* 1335.
- Connors, E., Lundregan, T., Miller, N., and McEwen, T. (1996) *Convicted by Juries, Exonerated by Science: Case Studies in the Use of DNA Evidence to Establish Innocence After Trial*, National Institute of Justice Research Report. Washington DC: US Department of Justice. [www.ncjrs.gov/pdffiles/dnaevid.pdf](http://www.ncjrs.gov/pdffiles/dnaevid.pdf).
- Cooke, G. (2007) ‘Are All Our Experts and Laboratories “Fit for Purpose”?’ Issue 10 *Archbold News* (7 December).



- Cowell, R.G., Lauritzen S.L., and Mortera, J. (2011) 'Probabilistic Expert Systems for Handling Artifacts in Complex DNA mixtures' 5 *Forensic Science International: Genetics* 202.
- Curran, J., and Buckleton, J. (2011) 'An Investigation into the Performance of Methods for Adjusting for Sampling Uncertainty in DNA Likelihood Ratio Calculations' 5 *Forensic Science International: Genetics* 512.
- Donnelly, P., and Friedman, R.D. (1999) 'DNA Database Searches and the Legal Consumption of Scientific Evidence' 97 *Michigan Law Review* 931.
- EU Council (2009). *Resolution on the exchange of DNA analysis results*, 30 November 2009 (O.J. 2009, C 296/1).
- Evett, I.W. (1992) 'Evaluating DNA Profiles in the Case where the Defence is "It Was My Brother"' 32 *Journal of the Forensic Science Society* 5.
- Evett, I., and Pope, S. (2011) 'Simulation Experiments to Investigate the Probability Distribution for the Number of Matching Alleles When a Reference DNA Profile Is Compared to a Multiperson Crime Profile' FSS Report TN-927.
- Foreman, L., and Evett, I.W. (2001) 'Statistical Analyses to Support Forensic Interpretation for a New Ten-locus STR Profiling System' 114 *International Journal of Legal Medicine* 147.
- Forensic Science Service (2004) *Guide to DNA for Lawyers and Investigating Officers*. [www.cps.gov.uk/legal/assets/uploads/files/lawyers'%20dna%20guide%20kswilliams%20190208%20\(i\).pdf](http://www.cps.gov.uk/legal/assets/uploads/files/lawyers'%20dna%20guide%20kswilliams%20190208%20(i).pdf).
- Gill, P., Ivanov, P., Kimpton, C., Piercy, R., Benson, N., Tully, G., Evett, I., Hagelberg, E., and Sullivan, K. (1994) 'Identification of the Remains of the Romanov Family by DNA Analysis' 6 *Nature Genetics* 130.
- Gill, P., Puch-Solis, R., and Curran, J. (2009) 'The low-level-DNA Threshold – Its Determination Relative to the Risk Analysis for National DNA Databases' 3 *Forensic Science International: Genetics* 104.
- Gill, P., Whitaker, K., Flaxman, C., Brown, N. and Buckleton, J. (2000) 'An Investigation of the Rigor of Interpretation Rules for STRs Derived From Less Than 100 pg of DNA' 112 *Forensic Science International* 17.
- Kaye, D.H. (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.

- Kaye, D.H. and Sensabaugh, G. (2011) 'Reference Guide on DNA Identification Evidence' in Federal Judicial Center and National Research Council, *Reference Manual on Scientific Evidence*, 3rd edn. Washington, DC: National Academies Press.
- Koehler, J.J. (1996) 'On Conveying The Probative Value Of DNA Evidence: Frequencies, Likelihood Ratios, and Error Rates' 67 *University of Colorado Law Review* 859.
- Koehler, J.J. (2001) 'The Psychology of Numbers in the Courtroom: How to Make DNA-Match Statistics Seem Impressive or Insufficient' 74 *Southern California Law Review* 1275.
- Law Commission (2011) *Expert Evidence in Criminal Proceedings in England and Wales*. Law Com No. 325. [www.lawcom.gov.uk/expert\\_evidence.htm](http://www.lawcom.gov.uk/expert_evidence.htm).
- Lindsey, S., Hertwig R., and Gigerenzer, G. (2003) 'Communicating Statistical DNA Evidence' 43 *Jurimetrics* 147.
- Lynch, M. (2003) 'God's Signature: DNA Profiling, the New Gold Standard in Forensic Science' 27 *Endeavour* 93.
- Lynch, M., Cole, S.A., McNally, R., and Jordan, K. (2008) *Truth Machine: The Contentious History of DNA Fingerprinting*. Chicago: Chicago UP.
- McCartney, C. (2006) 'The DNA Expansion Programme and Criminal Investigation' 46 *British Journal of Criminology* 175.
- National Research Council (2009) *Strengthening Forensic Science in the United States: A Path Forward*. Washington, DC: National Academies Press.
- NPIA (2009). NDNAD annual report 2007-2009, National Police Improvement Agency. <http://www.npia.police.uk/en/14395.htm> (accessed 13 September 2011).
- NPIA (2010) Standards of performance and monitoring of suppliers of profiles to the National DNA database. CUSTP-GP-004, NDNAD Delivery Unit.
- Pattenden, R. (2008) 'Authenticating "Things" in English Law: Principles for Adducing Tangible Evidence in Common Law Jury Trials' 12 *International Journal of Evidence and Proof* 273.
- Pope S., Clayton T., Whitaker J., Lowe J., and Puch-Solis, R. (2009) 'More for the Same? Enhancing the Investigative Potential of Forensic DNA Databases' 2 *Forensic Science International: Genetics, Supplement Series* 458.

- Puch-Solis, R., Rodgers, L., Balding, D., Mazumder, A., Evett, I., Pope, S., and Curran, J. (2012) 'Evaluating Forensic DNA Profiles using Peak Heights and Allowing for Dropout and Stutter' *forthcoming*.
- Redmayne, M. (2001) *Expert Evidence and Criminal Justice*. Oxford: OUP.
- Redmayne, M., Roberts, P., Aitken, C. and Jackson, G. (2011) 'Forensic Science Evidence in Question' *Criminal Law Review* 347.
- 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.
- Saks, M.J., and Koehler, J.J. (2005) 'The Coming Paradigm Shift in Forensic Identification Science' 309 *Science* 892.
- Schweitzer, N.J., and Saks, M.J. (2007) 'The CSI Effect: Popular Fiction about Forensic Science Affects the Public's Expectations about Real Forensic Science' 47 *Jurimetrics Journal* 357.
- SWGDM (2010) *Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories*. [www.fbi.gov/about-us/lab/codis/swgdam-interpretation-guidelines](http://www.fbi.gov/about-us/lab/codis/swgdam-interpretation-guidelines).
- Taroni, F. and Aitken, C.G.G. (1998a) 'Probabilistic Reasoning in the Law, Part 1: Assessment of Probabilities and Explanation of the Value of DNA Evidence' 38 *Science and Justice* 165.
- Taroni, F. and Aitken, C.G.G. (1998b) 'Probabilistic Reasoning in the Law, Part 2: Assessment of Probabilities and Explanation of the Value of Trace Evidence other than DNA' 38 *Science and Justice* 179.
- The Fingerprint Inquiry: Scotland (2011) *The Fingerprint Inquiry Report*. Edinburgh. [www.thefingerprintinquiryscotland.org.uk/inquiry/3041.html](http://www.thefingerprintinquiryscotland.org.uk/inquiry/3041.html)
- Triggs, C.M., Harbison, S.A. and Buckleton, J. (2000) 'The Calculation of DNA Match Probabilities in Mixed Race Populations' 40 *Science and Justice* 33.
- Tucker, V., Hopwood, A., Sprecher, C., McLaren, R., Rabbach, D., Ensenberger, M., Thompson, J., and Storts, D. (2011) 'Developmental Validation of the PowerPlex1 ESI 16 and PowerPlex1 ESI 17 Systems: STR Multiplexes for the New European Standard' 5 *Forensic Science International: Genetics* 436.
- Tucker, V., Hopwood, A., Sprecher, C., McLaren, R., Rabbach, D., Ensenberger, M., Thompson, J., and Storts, D. (2012) 'Developmental Validation of the

PowerPlex1 ESX 16 and PowerPlex1 ESX 17 Systems' 6 *Forensic Science International: Genetics* 124.

Tvedebrink, T., Svante-Eriksen, P., Smidt-Mogensen, H., and Morling, M. (2009) 'Estimating the Probability of Allelic Dropout of STR Alleles in Forensic Genetics' 3 *Forensic Science International* 222.

Tvedebrink, T., Svante-Eriksen, P., Smidt-Mogensen, H., and Morling, M. (2012) 'Statistical Model for Degraded DNA Samples and Adjusted Probabilities for Allelic Dropout' 6 *Forensic Science International: Genetics* 97.

Whitaker, J., Cotton, E. and Gill, P. (2001) 'A Comparison of Characteristics of Profiles Produced with AMPFISTR SGMPlus Multiplex System for Both Standard and Low Copy Number (LCN) STR DNA Analysis' 123 *Forensic Science International* 215.

## Appendix B – The UK DNA Database and Familial Searching

B1 The National DNA Database (NDNAD) was set up in 1995, and has come to play a major role in criminal investigations and prosecutions in the UK (Bramley, 2009). The NDNAD enables investigators to establish associations between people and items or locations of interest (e.g. a weapon or a crime scene) through searching and matching DNA profiles stored in the database.

B2 ***Type of profiles stored in the NDNAD***

The NDNAD contains two different types of profiles: (i) reference profiles and (ii) questioned profiles.

B3 **(i) Reference profiles.** The NDNAD contains over 4.8 million reference profiles (NPIA, 2009). Under the Police and Criminal Evidence Act 1984 (PACE), as amended, the police have powers to take reference samples, usually a mouth swab, from detainees suspected of recordable offences. Once taken, samples are sent to a forensic service provider (FSP) for profiling and analysis. The FSP produces a full STR DNA profile and submits the result to be loaded onto the NDNAD.

Scotland and Northern Ireland also maintain their own DNA databases. Reference profiles taken within these jurisdictions are stored on local databases as well as being sent to the NDNAD. Databases are regulated in accordance with the law of the jurisdiction, and there are some important differences between these regimes. Under Scottish legislation,<sup>54</sup> for example, reference profiles must be removed from databases, including the NDNAD, if the suspect is acquitted. Currently, in the same situation in England and Wales such profiles would be retained. However, legislation now before Parliament<sup>55</sup> will narrow the circumstances in which the profiles of unconvicted individuals can be kept on the NDNAD, in the light of the ruling of the European Court of Human Rights in *S and Marper v UK*.<sup>56</sup>

---

<sup>54</sup> Criminal Procedure (Scotland) Act 1995 ss.18-19, as amended.

<sup>55</sup> Protection of Freedoms Bill 2010-12, Part I ([www.services.parliament.uk/bills/2010-12/protectionoffreedoms.html](http://www.services.parliament.uk/bills/2010-12/protectionoffreedoms.html)).

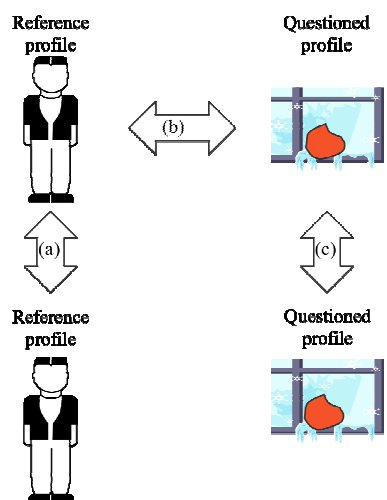
<sup>56</sup> *S v United Kingdom* (2009) 48 EHRR 5.

B4 (ii) **Questioned profiles.** The NDNAD holds about 350,000 questioned profiles produced from samples derived from criminal investigations. Some of these samples have a direct association with specific crimes, e.g. semen in a rape case. Other samples are collected for more general investigative purposes, e.g. cigarette ends from a scene of crime.

The police sometimes collect samples from volunteers for screening purposes, where a profile has been obtained from a crime sample and there is some information about the characteristics of the perpetrator, e.g. eyewitness accounts or CCTV footage. The people in the vicinity of the crime are asked to volunteer samples to eliminate themselves as potential suspects. These reference profiles are not stored in the NDNAD and are destroyed after being used for elimination purposes in the instant case.

B5 *Type of database searches*

There are three general types of database search, as depicted in Figure B1 (reference profiles indicated by people, and questioned profiles represented by blood samples):



**Fig. B1 Three types of NDNAD search.**

(a) Reference profile searched against reference profiles. The purpose is to identify and eliminate duplicates on the NDNAD. Individuals could be sampled more than once because they were arrested by different police forces at different times, or gave false names, or through spelling errors, etc.

(b) Questioned profile searched against reference profiles. The purpose is to identify individuals associated with crime scenes, and thus to produce potential suspects for further investigation.

(c) Questioned profile searched against questioned profiles. The purpose is to identify associated crime scenes which could be linked to a pattern of offending, e.g. a serial rapist or gang of armed robbers perpetrating multiple crimes.

B6 As DNA databases become larger, the possibility that *speculative searching* will produce ‘adventitious’ matches to innocent individuals grows. Thus, interpreting the results of speculative searching must be approached with care (Donnelly and Friedman, 1999).

B7 ***Familial Searching***  
Consider the situation where a standard DNA profile has been obtained from the scene of a crime, loaded and searched against the NDNAD, but no matching reference profile has been identified. One further investigative strategy, known as *familial searching*, is to assess whether any of the reference profiles in the database might be from a parent, child or full sibling of the unknown donor.

B8 **(i) Parent/child search:** Searches on the NDNAD for parent/child relationships are performed in three phases. First, the subset of reference profiles sharing at least one allele at every locus is selected. These profiles then are ranked according to their respective likelihood ratios, addressing the pair of hypotheses (propositions):

H<sub>1</sub>: the donor of the reference profile is the father or child of the donor of the questioned profile

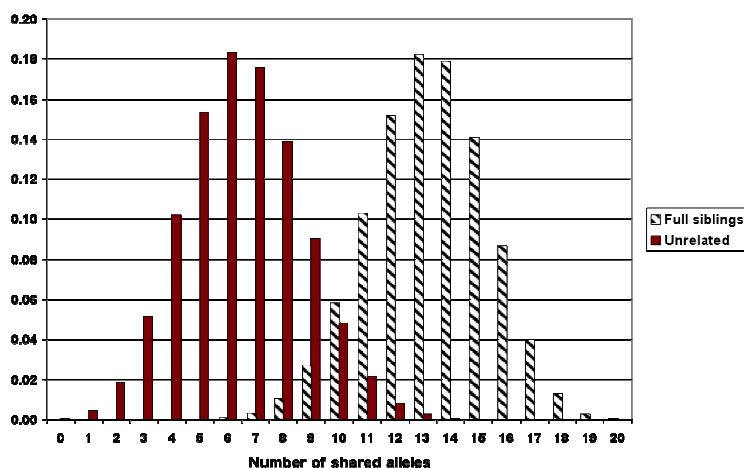
H<sub>2</sub>: the donors of the reference and questioned profiles are not blood related

In the third phase of the analysis, a further subset is selected from these profiles by considering other relevant circumstantial factors, such as geographical proximity, age and ethnic appearance of the offender (if known).

In one study reassessing five cases solved using parent/child searching, Pope et al. (2009) found that the parent (or child) of the offender was in the top fifty selected

reference profiles in three cases when ranked only by likelihood ratio. However, when other circumstantial factors were also taken into account, the parent (or child) was ranked first in four out of the five cases.

B9 (ii) **Sibling searches:** A similar searching methodology can be applied to siblings. Reference profiles are first ranked according to the number of alleles they share with the questioned profile. A subset of reference profiles sharing a specified number of alleles, usually 11 or 12 out of the possible 20 alleles of an SGMPlus profile, is then identified. This number has been selected as the optimum way to maximise the inclusion of full siblings while minimising the inclusion of unrelated individuals with overlapping genotypes, as depicted by Figure B2 and further elucidated by Evett and Pope (2011).



**Fig. B2 Computer simulation of probabilities of the number of shared alleles in a full SGMPlus profile for (i) unrelated individuals and (ii) full siblings (reproduced with the permission of the FSS)**

The subset of profiles is then further whittled down by likelihood ratio, addressing the pair of hypotheses (propositions):

H<sub>1</sub>: the donor of the reference profile is the full sibling of the donor of the questioned profile

H<sub>2</sub>: the donors of the reference and questioned profiles are not blood related

Finally, as before, other relevant circumstantial factors are taken into consideration. Pope et al. (2009) found that, in seven out of ten cases solved using full sibling searching, the sibling was in the top fifty reference profiles listed when ranked only



by likelihood ratio. When other factors were included, the sibling was ranked first or second in six of the seven cases.

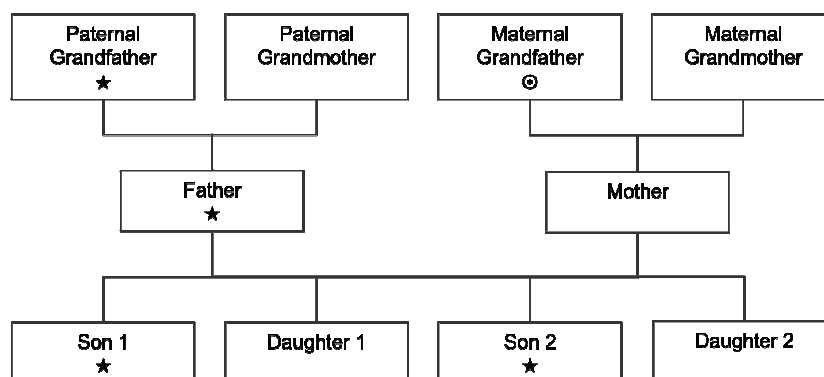
B10 For both parent/child and sibling familial searches, the most salient circumstantial factor was found to be geographical proximity between the questioned and reference samples. In other words, the offender's relatives tended to live close to the scene of the crime.

## Appendix C - Y-STR Profiles

### C1 Y-STR profiles

We saw in para.2.3 and Figure 2.2 that the sex test in standard DNA profiling comprises one locus on the X chromosome and one locus on the Y chromosome. However, there is another type of DNA test, focused exclusively on the Y chromosome, which profiles STRs at multiple loci.

C2 The Y-STR profiles of men in the same paternal line, i.e. men sharing a recent common male ancestor, are almost identical, because the rate of mutation of the Y-chromosome is relatively low. This pattern of genetic inheritance is depicted by Figure C1, in which the '★' chromosome passes down the male line from grandfather, to father, to sons.



**Fig. C1 Family tree of Y chromosome inheritance**  
(★ and ⊙ denote different Y-STR profiles)

However, the Y-STR profile of two men from different paternal lines could also be the same by chance. A Y-STR profile is therefore less discriminating than a standard DNA profile. Y-STR profiles, moreover, are not stored on the NDNAD, further reducing their forensic applications. Nonetheless, Y-STR profiles may be used in conjunction with other findings, including standard DNA profiles, for investigative purposes.

C3 There are many commercial systems available for conducting Y-STR tests using different loci. Figure C2 shows an example of a Y-STR profile produced using Y-filer™ for Applied Biosystems. This system targets 17 loci on the Y chromosome. There is just one allele per locus, since males have only one Y chromosome inherited from their fathers (whereas their non-sex chromosomes have two alleles at every locus, one from each parent).

**Fig. C2 - Y-STR profile**

DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS385a	DYS385b	DYS437	DYS438	DYS439	DYS448	DYS456	DYS458	DYS635	YGATAH4
14	12	28	24	10	13	13	11	14	15	13	11	20	15	16	24	12

C4 Y-STR profiles are often used in the investigation and prosecution of sexual offences. For example, consider a rape case where the attacker did not ejaculate. A vaginal sample has been taken from the victim and a DNA profile produced. Most of the DNA from the resulting mixed profile is related to the victim, with a small contribution from the attacker (registered on the EPG as a weak or partial profile). A Y-STR profile of the male component of the sample would contribute additional information in addressing whether the suspect is the source of the male DNA.

Y-STR profiles can also be employed to reduce a pool of potential suspects as part of a mass screening. The Y-STR profile obtained from a crime stain is compared to the Y-STR profiles of the suspects, or if not available, a male in the paternal line of each suspect. Suspects with non-matching Y-STR profiles can be eliminated from further inquiries. Any mutation on the Y chromosome is especially helpful in discriminating between potential suspects in these circumstances, since only the actual donor is likely to share the mutation identified in the crime sample (or vice versa).

A third investigative application of Y-STR profiles is to determine the minimum number of male donors in a mixed sample, and their relative contributions. This might be relevant, for example, in relation to alleged ‘gang rapes’ involving multiple assailants.

Finally, Y-STR profiles can be used to infer an offender's ethnic appearance by comparing the questioned profile with an ethnically stratified geographical database. However, this approach could give rise to ethical concerns, e.g. if it were used as a pretext to harass local communities.

C5 ***Expressing the probative value of Y-STR profiles***

The first stage in assessing the probative value of a Y-STR profile (e.g. a profile derived from scrapings taken from under the complainant's fingernails) is to consider the relevant suspect population. This could be all adult males in the locality (or further afield), or it might be possible to narrow down the suspect population in the light of other known information, e.g. eyewitness accounts.

In the absence of any further information about the offender's appearance or other identifying characteristics, data covering a range of populations can be considered.

C6 International reference databases are collated and managed in the Y-Haplotype Reference Database (YHRD), which is freely available through the internet ([www.yhrd.org](http://www.yhrd.org)). These data cover many of the populations of the world, including a set of British profiles produced using a variety of multiplexes.

In addition, the FSS holds a database of British Y-STR profiles classified into four ethnic appearance groups: Caucasian, Afro-Caribbean, Asian and East Asian (with 382, 525, 464 and 112 profiles, respectively). These data will still be available, as part of the YHRD, after the FSS ceases operational activities.

C7 The convention for reporting the significance of matching Y-STR profiles is to state the number of times that this profile occurs in the relevant database (its relative frequency). For example, a forensic scientist might report (or testify):

*In my opinion, the Y-STR profile from Mr X, which matches that of the crime stain, would be expected to be observed in fewer than 1 in 200 randomly chosen males in the British Caucasian population.*

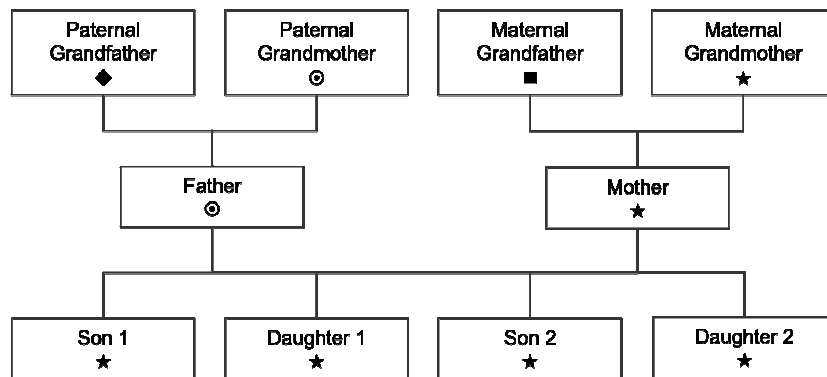
## Appendix D - Mitochondrial DNA Profiles

### D1 *mtDNA profiles*

As we noted in para 2.6, mitochondria are found in all cells of the human body. They are the ‘power plant’ providing energy for cellular processes. A cell contains thousands of copies of the small *mitochondrial DNA (mtDNA)* molecules, whereas there is only one set of chromosomes in the nucleus.

Being a much smaller molecule than a chromosome, mtDNA contains less information than nuclear DNA. Nonetheless, just like Y-STR profiles (which only contain information from the Y chromosome), mtDNA can still prove useful in forensic analyses because it is more abundant and resistant to degradation.

D2 The mtDNA profiles of men and women in the same maternal line, i.e. people sharing a recent common female ancestor, would be almost identical, because the rate of mutation of mtDNA is relatively low. This pattern of genetic inheritance is depicted by Figure D1.



**Fig. D1 Family tree showing mtDNA inheritance**  
(◆, ◎, ■, ★ denote different mtDNA profiles)

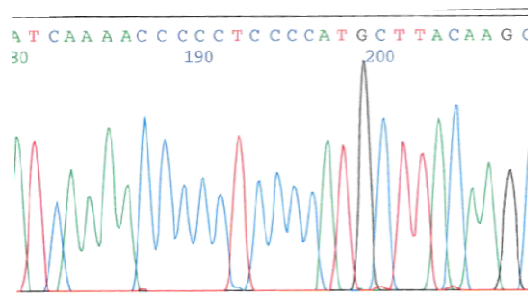
As with Y-STR profiles, however, the mtDNA profiles of two unrelated people with different maternal lines could be the same by chance.

D3 Males in the same paternal line share the same Y-STR profile, but females have no Y chromosome at all and therefore cannot be identified by Y-STR profiling. By contrast, males and females alike inherit mitochondria from their mothers.

An mtDNA profile is less discriminating than a standard (nuclear) DNA profile. It cannot be compared with Y-STR profiles, nor are mtDNA profiles stored on the NDNAD, further reducing their investigative applications. However, used in conjunction with other findings (including standard DNA profiles), mtDNA profiles can contribute to criminal investigations and prosecutions.

D4 Paralleling the process in nuclear DNA profiling, an mtDNA profile is obtained by determining the sequence of base pairs in the DNA molecule. A questioned profile can then be compared to a reference profile in the usual way. In contrast to standard DNA and Y-STR profiles, however, it is not possible to determine the number of donors or their respective contributions to a mixed sample.

Figure D2 reproduces a section of a sequence of about 30 bases in an mtDNA profile. The full profile would consist of about 800 such bases.



**Fig. D2 Illustration of a sequence from an mtDNA profile**

D5 The most common forensic application of mtDNA profiles relates to hairs recovered from crime scenes. If the hair comes with its root intact, it is usually possible to obtain a standard DNA profile. However, if the hair has no root, precluding nuclear DNA profiling, an mtDNA profile might still be obtained from the shaft.

Like Y-STR profiles, mtDNA can also be used for investigative purposes to infer the ethnicity of a person (with the attendant ethical concerns).

Other common applications include identifying the bodies of missing persons and disaster victim identification (DVI) work. mtDNA does not degrade as quickly as nuclear DNA where human remains are skeletonised or badly burnt. Family members in the same maternal line can be profiled for comparison. For example, Prince Philip provided a sample for mtDNA testing which helped identify the bones of Romanov family members (Gill et al., 1994).

D6 *Expressing the probative value of mtDNA profiles*

The first stage in assessing the probative value of an mtDNA profile (e.g., derived from a hair) is to consider the relevant donor population. As in relation to Y-STR profiles, the relevant population depends on geographical factors and other circumstantial information (if any) known at the time. For example, is the donor a vagrant pulled out of the local canal or the victim of a plane crash?

In the absence of case-specific information, international databases of mtDNA profiles, which are available in print and electronically, may be consulted in order to try to infer the donor's ethnic appearance. One widely used database, EMPOP ([www.empop.org](http://www.empop.org)), contains the profiles of about 4,500 individuals, mostly of European origin.

D7 The conventional method of reporting the significance of matching profiles is to state the number of times that this profile has been seen in the relevant, ethnically stratified database (i.e. its relative frequency). For example, in the context of criminal proceedings in which Mr X is the suspected donor of crime stain mtDNA, an expert might testify:

*In my opinion, the mtDNA sequence from Mr X, which matches that of the crime stain, would be expected to be observed in fewer than 1 in 3,000 randomly chosen people in the European Caucasian population.*

The form of the evidence resembles the way in which serology testimony used to be given under the old ABO blood-typing system (itself rendered virtually obsolete by forensic DNA profiling), albeit with smaller – and therefore, generally speaking, more probative - relative frequencies.

## Appendix E – Glossary

**Adventitious match:** matching a questioned *DNA profile* to the profile of an individual who is not the source of the crime stain DNA (e.g. through *speculative searching* of a DNA database).

**Aliquot:** a sample taken from extracted DNA.

**Allele:** a variation at a given *locus* on a *chromosome*. Alleles occur in pairs, one on each chromosome. In the context of forensic DNA profiling, an allele is the number of *short tandem repeats (STRs)* at a locus in a chromosome.

**Allele probability:** an estimate of the probability of a *genotype* in a designated population.

**Allelic drop-out:** absence or non-detection of one or more alleles from a DNA profile.

**Amplification:** the process by which the number of copies of specific DNA sequences are increased using a sequential copying process.

**Autosomal DNA:** any *chromosome* other than a sex chromosome.

**Base pairs:** DNA is formed from four chemical ‘bases’ called adenine, cytosine, guanine and thymine, usually represented by the letters *a*, *c*, *g* and *t*. A base pair is a base in one strand of the *double helix* together with its complementary base in the other strand.

**Body fluid:** in forensic contexts, specific tissue-types including blood, semen, saliva, hair, and epithelial (e.g. skin) cells.

**Cell:** the microscopic ‘building block’ and smallest working unit of an organism.



**Chromosomal DNA** (or *nuclear DNA*): DNA in the *chromosomes*, which are located in the nucleus of a cell.

**Chromosomes:** DNA molecules. Humans have 23 pairs of chromosomes (including one pair of sex chromosomes).

**Co-ancestry allowance:** an adjustment made in calculating genotype probabilities to allow for possible recent shared ancestry of people of similar ethnic appearance.

**Co-ancestry coefficient;** or *population sub-structure correction*:. the allowance for possible shared ancestry within a population, expressed as a value between zero and one (higher values corresponding to greater shared ancestry).

**Cold hit:** identification of a suspect in a ‘cold’ case by searching *crime samples* against *reference samples* on a DNA database.

**Crime samples,** crime stains, questioned samples: samples of genetic material of unknown origin derived from (suspected) crime-scenes.

**Custodian of the UK National DNA Database (NDNAD):** responsible for ensuring the quality and standards of forensic suppliers approved to upload DNA profiles to the NDNAD.

**Degradation:** natural process by which the DNA molecule breaks down.

**DNA:** acronym of deoxyribonucleic acid.

**DNA (profiling) evidence:** relevant information prepared for the purposes of criminal proceedings, comprising the results of DNA profiling and their interpretation by a qualified forensic scientist.

**DNA profile:** the combined *genotypes* for all *loci* for an individual person. A full profile contains information at all the loci tested. A partial profile lacks information at one or more loci, typically the longer length STRs.

**DNA profile EPG:** electropherogram of a profile displaying peaks for each of the *alleles* in a *multiplex*.

**DNA profiling protocol:** the list of equipment and settings used to produce a *DNA profile*.

**DNA profiling system:** combination of the multiplex used to produce the DNA profile and a protocol for all the stages in the analytical and interpretation process.

**Double helix:** the characteristic, double-stranded form of the DNA molecule.

**EPG, electropherogram:** instrumental output showing DNA profile in the form of peaks on a graph.

**Gene:** the basic unit of heredity, composed of a sequence of DNA *base pairs* occupying a specific position on a chromosome.

**Genome:** the entire genetic material of an organism, contained in a full set of *chromosomes*.

**Genotype:** a pair of alleles at a designated locus, or collection of pairs of alleles across multiple loci. DNA profiling distinguished between different genotypes.

**Heterozygous, heterozygote:** genotype at a locus with two different length STRs, e.g. 17,18.

**Heterozygote balance:** the degree of balance, or imbalance, between two peaks at a locus on an **EPG**.

**Homozygous, homozygote:** genotype at a locus with two STRs the same length, e.g. 18,18.

**Inhibition:** chemical reaction causing the DNA *amplification* process to be sub-optimal or to fail altogether.

**Interpretation:** in relation to the results of DNA profiling, interpretation covers (i) designating the alleles in a DNA profile; (ii) the resolution of a mixture; and (iii) calculating a *likelihood ratio*.

**LCN, low copy number:** a form of analysis applied to **LTDNA samples**, more sensitive than standard profiling process and often involving increased *amplification* cycles.

**Likelihood ratio, LR:** the ratio of the probability of an event assuming one set of circumstances to be true, to the probability of the same event assuming another set of circumstances to be true; e.g. the probability of the evidence assuming the prosecution's proposition to be true and the probability of the evidence assuming the defence proposition to be true.

**Loci, sing. locus:** regions of *non-coding DNA* consisting of *short tandem repeats* (**STRs**) which vary in length from person to person.

**LTDNA, samples:** DNA present in small amounts, exhibiting *stochastic variation*.

**Match probability:** see *random match probability*.

**Mitochondria, sing. mitochondrion:** Organelles (parts of cells) containing their own DNA. Several thousand are present in each cell, in contrast to 22 *chromosomes* in the nucleus of the cell.

**Mitochondrial DNA:** DNA contained in *mitochondria*, which may be recovered for use in forensic DNA profiling. It is inherited only through the maternal line.

**Mitochondrial DNA profile:** profile obtained from *mitochondrial DNA*.

**Mixture, DNA mixture:** DNA deriving from more than one individual extracted from the same stain (which could also contain different body fluids).

**Multiplex:** the combination of *loci* analysed in a particular *DNA profiling system*.

**Mutation:** deviation in the DNA sequence owing to the failure of cellular DNA to reproduce itself accurately. Mutations might involve adding, changing or deleting one or more base pairs at particular loci. Genetic mutations can have serious implications for an individual's health and wellbeing, but they are irrelevant for most forensic applications, which concentrate on *non-coding regions*.

**Nanogram:**  $10^{-9}$  grams, or 1000 picograms.

**National DNA Database (NDNAD):** the UK's national DNA database, created in 1995 to store DNA profiles taken from convicted offenders, individuals arrested on suspicion of a recordable offence and unsolved crime scenes. The database facilitates *speculative searching* possibly leading to a *cold hit*, but also running the risk of *adventitious matches*.

**Non-coding (DNA) regions:** those regions of the DNA molecule particularly suitable for forensic profiling, owing to the extent of detectable variation between individuals in those regions.

**Nuclear DNA:** a synonym for *chromosomal DNA*.

**Picogram:**  $10^{-12}$  grams, or 0.001 nanograms.

**Polymerase Chain Reaction, PCR:** technical name for the DNA *amplification* process.

**Population sub-structure correction:** see *co-ancestry co-efficient*.

**Process control samples:** dummy samples run to check that the profiling process is secure from contamination. They could be (i) negative or 'reagent blank'

controls containing no DNA at all; or (ii) positive controls containing a specified quantity of a known profile, which should replicate expected results.

**Proposition:** a statement with a truth-value, asserting that  $x$  is – or alternatively, that  $x$  is not – the case. The *likelihood ratio* approach to assessing the probative value of evidence considers the probability of the evidence under two mutually exclusive propositions, ‘the prosecution proposition’ (e.g. that the accused is the source of the crime stain profile) and ‘the defence proposition’ (e.g. that a person not blood related to the accused is the source of the crime stain profile).

**Prosecutor’s fallacy**, illegitimate transposition of the conditional: treating the probability of the evidence (e.g. a DNA match), assuming the defence proposition (e.g. that the donor of the crime stain DNA is not the accused), as though it were the probability of the defence proposition in light of the evidence (e.g. the probability that the accused is not the donor of the crime stain DNA in light of a matching profile).

**Protocol:** the combined validated methods used to obtain, analyse and interpret DNA profiles. Also see *DNA profiling protocol*.

**Quantification:** measurement of the amount of DNA in a sample or extract.

**Questioned sample:** a sample taken from a crime scene or from people or objects whose origin is at issue in criminal proceedings.

**Random match probability, RMP:** the probability of observing a genotype of an unknown person that is the same as the profile from a crime scene stain, given knowledge that the accused’s genotype also matches the crime stain profile, and assuming that the unknown person and the accused are neither the same person nor a blood relative.

**Random occurrence ratio:** a (disfavoured) synonym for the RMP coined by the Court of Appeal in *R v Doheny and Adams* [1997] 1 Cr App R 369, CA.

**Reference DNA profile:** A DNA profile produced from an optimal amount of DNA from a *reference (DNA) sample*.

**Reference (DNA) sample:** physical sample containing DNA (e.g. from a buccal swab) taken from a known source under controlled conditions for the purposes of forensic profiling.

**Relative fluorescence unit(s) (rfu):** unit of measurement of peak heights on an *electropherogram* (EPG).

**Relative frequency:** the number of times that an event of interest (e.g. a DNA profile) occurs in – *relative to* – a designated population. Thus, the relative frequency of  $x$  (events) in  $y$  (population) is the number, or percentage, of  $y$ s that are also  $x$ s.

**Sex chromosomes:** X and Y *chromosomes* in humans. Males have one X and one Y chromosome, females have two X chromosomes.

**Sex test:** the part of a *multiplex* which detects the presence of X and Y chromosomes, indicating the possible sex of the donor of the DNA.

**SGMPlus, Second Generation Multiplex Plus:** the multiplex currently required for loading a DNA profile on to the NDNAD. It consists of 10 **STRs** and a *sex test*.

**Speculative searching:** comparing a crime sample against reference profiles on a DNA database, which may produce a *cold hit*.

**Stochastic variation:** random variation. In relation to DNA profiling, variation in the subsets of alleles that could be selected on repeat sampling from DNA with small numbers of copies (e.g. if alleles A and B are present, sampling might select A and B, A only, B only or neither).

**Stochastic threshold:**

(1) the minimum quantity of DNA needed to produce a reliable profile. The Court of Appeal in *R v Reed and Reed* [2010] 1 Cr App R 23; [2009] EWCA

Crim 2698, [74], observed that ‘[t]here is no agreement among scientists as to the precise line where the stochastic threshold should be drawn, but it is between 100 and 200 picograms’.

(2) the peak height value, in rfu units, above which it is reasonable to assume that, at a given locus, *allelic dropout* of one of the two alleles of a heterozygote has not occurred.

**STRs**, short tandem repeats: short sequences of *base pairs* in a DNA molecule that are repeated multiple times. They typically consist of four, or sometimes three, base pairs.

**Stutter**: an artefactual peak sometimes seen in an EPG, one STR unit to the left of a genuine allelic peak.

**Stutter ratio**: ratio of the height of a *stutter* peak to the height of its parent allelic peak.

**Transposing the conditional**, illegitimately: see *prosecutor’s fallacy*.

**Validation**: the process by which a new forensic method or technique is assessed to ensure that it is fit for purpose in the administration of criminal justice, and will continue to function properly once implemented. Regulatory requirements may require compliance with external quality standards (e.g. ISO 17025: General Requirements for the Competence of Testing and Calibration Laboratories).

**Y-STR profile**: A profile obtained from the Y chromosome, which is present only in males. Since all the *loci* tested are on the same chromosome, the profile consists of only one allele at each locus.







