Guide to DNA for Lawyers and Investigating Officers

This booklet is designed to give lawyers and investigating officers a basic understanding of DNA analysis and interpretation. It aims to assist them in the investigation of criminal offences involving DNA evidence and when dealing with expert witness evidence in court.
# Contents:

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>5</td>
</tr>
<tr>
<td>What is DNA?</td>
<td>6</td>
</tr>
<tr>
<td>Definition</td>
<td>6</td>
</tr>
<tr>
<td>Nuclear DNA</td>
<td>6</td>
</tr>
<tr>
<td>Mitochondrial DNA</td>
<td>7</td>
</tr>
<tr>
<td>How is DNA profiling done?</td>
<td>8</td>
</tr>
<tr>
<td>Types of DNA profile</td>
<td>10</td>
</tr>
<tr>
<td>Nuclear DNA</td>
<td>10</td>
</tr>
<tr>
<td>Y-STR DNA profiles</td>
<td>12</td>
</tr>
<tr>
<td>Mitochondrial DNA</td>
<td>13</td>
</tr>
<tr>
<td>Dealing with issues of contamination</td>
<td>15</td>
</tr>
<tr>
<td>The National DNA Database® and the role of the custodian</td>
<td>17</td>
</tr>
<tr>
<td>Familial searching</td>
<td>20</td>
</tr>
<tr>
<td>Principles for the evaluation of DNA evidence</td>
<td>21</td>
</tr>
<tr>
<td>Assessing the value of the DNA evidence</td>
<td>21</td>
</tr>
<tr>
<td>Match probability of one in a billion</td>
<td>23</td>
</tr>
<tr>
<td>Reporting a DNA match to court.</td>
<td>24</td>
</tr>
<tr>
<td>Uniqueness probability</td>
<td>25</td>
</tr>
<tr>
<td>Scientist's statement</td>
<td>26</td>
</tr>
<tr>
<td>The prosecutor's fallacy</td>
<td>27</td>
</tr>
<tr>
<td>Hierarchy of propositions</td>
<td>28</td>
</tr>
<tr>
<td>The evaluation of DNA evidence in database matches</td>
<td>30</td>
</tr>
<tr>
<td>Presentation of evidence and statements</td>
<td>32</td>
</tr>
<tr>
<td>Further advice and suggested reading</td>
<td>33</td>
</tr>
<tr>
<td>Appendix</td>
<td>34</td>
</tr>
<tr>
<td>Glossary of terms and abbreviations</td>
<td></td>
</tr>
</tbody>
</table>
Introduction

This booklet is designed to give investigating officers and lawyers a basic understanding of DNA analysis and interpretation to assist them in the investigation of criminal offences involving DNA evidence and when dealing with expert witness evidence in court.

This is an updated version of the ‘Lawyers’ Guide to DNA’, first published in March 1998, which has been rewritten to include new topics brought about by advances in DNA profiling.

The original version is still available from The Forensic Science Service® (FSS®) and as many of the principles are still relevant should be referred to if the reader has little or no previous knowledge of DNA. For a glossary of terms used in this booklet see the appendix on page 33.
What is DNA?

DNA stands for Deoxyribonucleic Acid.

This is the name of the complex chemical found in virtually every cell in the body. It is like a very long, miniature piece of string containing the coded sequence that determines our physical characteristics and directs all the chemical processes in the body.

DNA is found in two different parts of each cell - the nucleus and the mitochondria.

Nuclear DNA

Almost all cells in the body contain a nucleus. The genetic information coded for by nuclear DNA is carried in the chromosomes from one generation to the next, half of a person’s nuclear DNA being inherited from the mother and half from the father. The nuclear DNA of identical twins is expected to be the same. Other siblings inherit different combinations of nuclear DNA from the same parents and their DNA is therefore somewhat different from one another. The DNA from unrelated individuals is likely to be even more different. Each generation of people is a new and different combination of genetic material from the previous generation.
If it were possible to examine the entire DNA molecule from every person in the world it would be possible to ascribe, using DNA technology alone, a sample of DNA to one specific person (except identical twins). However, it is doubtful that this would ever be practicable.

Instead, the techniques used by the FSS look only at a set of components of the DNA molecule. Each one of these components is known as a locus (plural loci). The loci have been selected because they are regions where there is known to be considerable variation between people caused by short pieces of DNA code being repeated over and over again, end to end. They are called short tandem repeat loci (STRs). Although any given set of components (the DNA profile) will not be unique to an individual, methods have been developed to calculate the probability of one person's DNA profile matching that of another just by chance.

It is thus not possible, in general, to express the kind of categorical opinions of identity of source that are presented from fingerprint evidence. Nevertheless, DNA profiling is highly discriminating and can provide extremely powerful evidence for presentation at court.

**Mitochondrial DNA**

Mitochondrial DNA (mtDNA) is passed from mother to child with no contribution from the father. Each sibling will have the same mtDNA type and this will be passed on through the maternal line for many generations without significant change. The analysis and interpretation of mtDNA is different from that for nuclear DNA.

This booklet deals primarily with the analysis and interpretation of profiles obtained from nuclear DNA, (see page 13) has a brief outline of mtDNA analysis.
How is DNA profiling done?

DNA profiling has developed rapidly in forensic science since the late 1980s. Great advances have been made in the automation and computerisation of the technique, and there have also been improvements in the sensitivity and application of the method. Today, the FSS is recognised as a world leader in DNA profiling techniques, using advanced equipment and methods.

The main steps in the technical process are:

- **Extraction**
  The DNA is isolated and separated from other cellular material.

- **Quantification**
  The amount of DNA extracted from the sample is measured to determine the optimal amount of DNA required for the next step of the process.

- **Amplification (Polymerase Chain Reaction - PCR)**
  Specific areas of interest (loci) within the DNA molecule are targeted and multiple copies produced.

- **Electrophoresis**
  The amplified DNA is separated according to size.

- **Analysis**
  The DNA is measured in such a way as to allow convenient comparison with other DNA profiles and storage of the profile on The National DNA Database®.
STR analysis is the standard method used for both intelligence and evidential purposes. The STRs have no function in passing on physical characteristics - indeed, the precise function of the STR loci is still a subject of debate among scientists.

From a forensic point of view, the interesting feature of an STR locus is that it consists of two alleles, one inherited from the mother and one from the father, and the number of repeats for each allele can vary independently of each other. Each allele has a name that reflects the number of repeats. “Allele 12” would indicate the presence of 12 repeats, “allele 14” would consist of 14 repeats, and so on.

So, at one locus, a person may be designated (12,14) for example - indicating that he/she has one allele of 12 repeats at that locus and another of 14 repeats. It is, of course, possible for a person to inherit an allele with the same number of repeats for a given locus from both parents, so a person who has inherited, for example, the allele with 11 repeats from both parents would have a locus designated (11,11).
Types of DNA Profile

Nuclear DNA

SGM / FSS SGM Plus profiles

The second generation multiplex (SGM) system was introduced by the FSS in 1995 for DNA intelligence database purposes and in 1996 for the analysis of biological materials from a crime scene. The system looked at six STR loci and also tested for the gender of an individual. The probability that two unrelated people would have the same SGM profile is of the order 1 in 50 million.

The SGM Plus system was introduced in 1999. This tests for the same six STR loci as SGM, along with the gender marker and four new STRs. This makes it far more discriminating than SGM. It is also is more sensitive due to improvements made in the process. The FSS can now analyse samples that would not previously have been successfully analysed with SGM because of the increased sensitivity.

Due to the overlap of six of the STR loci the two systems are compatible. Therefore, samples processed using SGM Plus can still be compared with samples previously loaded to The National DNA Database® that had been analysed using the SGM system. The SGM profiles can also be upgraded to SGM Plus to improve the level of discrimination and reduce the risk of “adventitious matches”. An adventitious match is obtained when the DNA profiles from two individuals match just by chance.

To illustrate the process, consider this hypothetical case. A criminal justice (CJ) sample, taken from a cheek scrape or hair root, was analysed with SGM in 1997 and its profile was subsequently loaded onto The National DNA Database®. In 1999, a crime stain was processed with SGM Plus and, when its profile was searched against The National DNA Database®, it matched that of the original CJ sample. The individual concerned, when questioned about the later offence, professed his innocence, stating that he had never even been to the location of the crime. The
FSS can “upgrade” the original profile using SGM Plus. If the profiles still match, the strength of the evidence against him is increased. If they no longer match then he can be eliminated from being the source of the DNA.

**Full / Partial DNA profiles**

A *full* SGM Plus DNA profile is obtained when the analysis of all ten STR loci and the gender marker has given successful results. A full profile would be expected from items that are not degraded, such as reference blood samples, cheek scrapes and fresh samples from the scene of the crime. A *partial* DNA profile occurs when it has not been possible to obtain results for one or more of the STR loci or the gender marker. This can happen when the sample being analysed has degraded after being exposed to the environment for a period of time, and/or when there is very little material available.

Partial DNA profiles obtained from a crime sample can be held on and searched against The National DNA Database as long as they meet certain minimum criteria. The more partial the profile, however, the greater the number of matches that will result and the lower their intelligence value. Partial profiles that have only a few components of the DNA present and do not meet the minimum loading criteria can still be searched against The National DNA Database by carrying out ‘one-off’ speculative searches. The intelligence value in such searches depends on what other information might be available to help restrict the large number of matches likely to be obtained.

**DNA mixtures**

A mixed profile will result when DNA from two or more individuals is present in a sample. This can often happen, for example, in sexual offences and in violent crime where the attacker *and* the victim have both contributed.

Where semen has been recovered from rape victims it is sometimes possible to physically separate the male and female components of the mixture and thus obtain the individual DNA profiles.
Where physical separation is not possible, identification of the possible profiles of each contributor may still be possible, depending on the relative proportions of DNA from each that is present. This is simplified where reference samples are available from at least one of the contributors to the mixture.

For two-component mixtures, where the relative proportions are somewhat similar, the FSS has developed a technique - pendulum list search (PLS) - for identifying and prioritising the possible combinations of individual profiles that could have made up the mixed profile.

**LCN (low copy number) DNA profiles**

Low copy number (LCN) DNA profiling has been developed through extensive research by the FSS. It exploits the increased sensitivity of the SGM Plus system and enables the successful analysis of samples that would previously not have been expected to yield results (e.g. the cellular residue in a fingerprint or on a swab from the inside of a watch strap). LCN DNA profiling has enabled the FSS to analyse samples from ‘cold’ cases from which it had not previously been possible to obtain a DNA profile. These samples have often been stored for many years in the hope that improvements in forensic techniques will allow these cases to be revisited at a later date.

LCN DNA profiles are more likely to be partial and/or mixtures. It may also not be possible to indicate the type of fluid that the profile came from. It is, therefore, important with LCN DNA profiles that discussions are held between the investigators, lawyers and the scientist to assist with the evaluation of the findings.

**Y-STR DNA profiles**

The male Y chromosome in nuclear DNA also contains STRs. These Y-STRs are passed down the male line, from fathers to sons without change over many generations. They are different from those used for SGM/SGM Plus profiling but they can be used to produce Y-STR profiles.
Y-STR profiling is helpful in comparing the male components of mixtures containing male and female DNA and for the identification of missing relatives in the male line. Because it runs unchanged through generations it can also provide indications of ethnic/geographic origins.

Y-STR profiles can be useful in the investigation of families identified by the technique of familial searching (see page 20). This will enable police to maximise the use of collected DNA without causing further public intrusion.

Mitochondrial DNA

Mitochondria are found in every cell of the body and provide the energy that is required for cellular processes. They have their own DNA, which is different from the nuclear DNA and contains much less information. However it is far more abundant within each cell. Partly due to this abundance, mitochondrial DNA (mtDNA) analysis is very sensitive and can be used when the conventional STR techniques (using nuclear DNA) would not normally result in a full profile (e.g. when analysing shed hairs, aged bones and teeth).

The technique used for mtDNA analysis detects differences in the individual basic building blocks of the mtDNA molecule and is known as “mtDNA sequencing”. MtDNA sequencing is far less discriminating than nuclear STR DNA analysis but nevertheless is very useful when taken in the context of other evidence.

As mtDNA is maternally inherited, and all individuals who are related by a maternal link will have the same mtDNA profile, mtDNA profiling is also ideal for the identification of bodies (e.g. after mass disasters), where remains can be badly burnt or degraded, using samples from maternal relatives for comparison.

Because mtDNA is relatively stable over generations it can also be used to provide indications of ethnicity.
Some of the disadvantages of mtDNA analysis include:

- All maternal relatives being expected to have the same mtDNA profile, complicating the interpretation of a match between a suspect and crime sample.

- MtDNA profiles being incompatible with STR profiles and not being able to be checked against The National DNA Database®.

- MtDNA analysis not being amenable to the analysis of mixed profiles, so that only items that are known to have originated from one individual (e.g. bone, hair and faeces) are suitable to be processed.

In a similar vein to Y-STR’s, mtDNA can be used in the investigation of families by looking at the female line.
Dealing with the issues of contamination

Any contact between a suspect or victim and an object or scene prior to a crime being committed can result in DNA being transferred and later recovered. This is not, strictly speaking, contamination, but it is, of course, an important consideration that has to be taken into account.

The word “contamination” is used to denote the presence of any DNA in a sample that has been introduced post-incident. This can be during examination of the scene or during collection or analysis of the sample, either from the individuals involved, from contact with DNA from other items, from the materials used for sample collection or analysis, or from the surrounding environment. It is not associated with the offender or victim in question.

The FSS takes enormous care in the analysis of samples in the laboratory to ensure the integrity of DNA profiles obtained from crime and reference samples. This includes: the use of protective clothing by staff; very strict protocols for the handling and processing of samples, and environmental control, in line with international quality assurance guidelines; monitoring of the materials used to carry out analyses to ensure that they are DNA-free; processing of DNA-free blanks; checking for contamination between samples during analysis and for contamination by staff; carrying out a proportion of analyses in duplicate; and monitoring of the laboratory environment and equipment.

This allows the FSS to minimise the risk of contamination and to take appropriate corrective action when it does occur. This action may include re-analysing the sample, if appropriate, or taking the contamination into account when providing intelligence information to the investigator or presenting evidence in court.
A similar degree of care must also be taken at the scene, where the risk of contamination is possibly greater. Detectable levels of DNA can be deposited by investigators and those involved in the recovery of evidence. By following best practice in crime scene management (i.e. wearing gloves, facemasks, scene suits and mob caps) the risk of compromising potential evidence can be greatly reduced. The FSS is committed to helping manage the contamination risk by providing quality controlled sampling kits and keeping staff and investigators aware of anti-contamination practices through the use of literature updates, training courses and the *Scenes of Crime Handbook*.

The identification of contamination by individuals in the investigative and analytical processes is facilitated by the availability of elimination databases. One of these, the police elimination database (PED), contains profiles of those police employees who are likely to visit a crime scene or deal with the victim/suspect. Profiles from named individuals on the PED are checked against a specific crime sample profile when it is considered that contamination by that individual could have occurred. Another contains profiles of FSS staff who may come into contact with casework items or the facilities used for examination or analysis. All profiles generated by the FSS are routinely checked against this staff elimination database (SED).
The National DNA Database® and the role of the custodian

The National DNA Database® was set up in 1995, following amendments to the Police and Criminal Evidence Act (PACE) 1984 by the Criminal Justice and Public Order Act 1994. These amendments allowed CJ samples (cheek scrapes or rooted hairs) to be obtained for DNA profiling, in broadly the same circumstances as fingerprints, and for the profiles obtained from these to be searched against records held by or on behalf of the police. Further amendments to PACE were introduced by the Criminal Evidence Act 1997, the Criminal Justice and Police Act 2001 and the Criminal Justice Act 2003. These extended the range of people from whom samples could be taken, allowed for people to volunteer to have their profiles added to The National DNA Database®, and removed the constraints on retention of samples and profiles when an individual was not prosecuted or was acquitted.
The FSS was appointed custodian of The National DNA Database® by the Association of Chief Police Officers (ACPO) for the first five years of its operation and this appointment has since been renewed. Under the terms of a memorandum of understanding between ACPO and the FSS, the FSS provides the IT facilities for holding The National DNA Database® and develops the software applications and procedures for generating the intelligence information required. The actual profile data remains the property of the individual police forces that obtained the samples for analysis.

The operation of The National DNA Database® is overseen by The National DNA Database Board, which meets quarterly and is chaired by the chief constable nominated by ACPO to lead on forensic science matters.

Other law enforcement agencies outside the remit of ACPO (e.g. in Scotland, Isle of Man and Jersey) have also agreed to abide by the memorandum of understanding and contribute samples for profiling for The National DNA Database®.

The FSS exercises the duties of custodian through the office of its chief scientist. As custodian, the chief scientist is accountable to The National DNA Database Board for maintaining standards of integrity and management of The National DNA Database® and providing the range and levels of services specified by the board.

The custodian sets the expected “Standards of Performance” for forensic science laboratories that wish to provide DNA profiles to The National DNA Database® and ensures that these standards are achieved and maintained. In addition to the FSS, there are now a number of other laboratories, in both the public and private sectors, that do this work. All use a compatible technique based on the analysis of STR regions of DNA using SGM Plus technology. The suppliers meet regularly under the chairmanship of the custodian to discuss technical issues and new developments.
CJ samples are classified as non-intimate samples and can be taken by the police from anyone arrested and in police detention, charged, about to be reported for, cautioned or convicted of a recordable offence. The mouth and hair samples are submitted by the police to an approved forensic science laboratory and the laboratories submit the profiles from these to the custodian for addition to The National DNA Database®.

The police also obtain samples from biological material left at scenes of crime. These samples are submitted for analysis to an approved forensic science laboratory and the profiles obtained are also then passed to the custodian for loading to The National DNA Database®.

As each new profile is added to The National DNA Database® it is routinely checked for matches against all other profiles held on the database. Such matches are reported as intelligence information to the police, linking a named individual with an unsolved crime or linking different unsolved crimes together. The intelligence matches may also be used as the basis for charging a suspect, so long as there is sufficient other supporting information. If, on the basis of the DNA match and other information implicating the individual with the offence, a prosecution is intended, the police are required to obtain a new sample and have this analysed specifically for use in evidence.

A match between two full SGM Plus profiles from unrelated individuals has a discriminating power of about one in one billion. Matches involving partial profiles or relatives are more likely to occur by chance and the discriminating power is thus much less.

The National DNA Database® was the first of its kind in the world and has received widespread acclaim as the most important advance since fingerprinting in the prevention, detection and deterrence of crime. It is a dynamic database, as profiles are constantly added to and removed from it.

Through the DNA expansion programme, managed by the Home Office, the government has invested considerable additional funding - £182.6m to the end of 2003/04 - to enable the police to increase their use
of DNA and through so doing to expand The National DNA Database®, so that it includes profiles from all active criminals. The funding has also included provision for developments that improve the efficiency and effectiveness of The National DNA Database®.

In May 2004, The National DNA Database® held around 2.3 million profiles from different individuals and around 229,000 profiles from scenes of crime samples. Around 35,000 new subject sample profiles and 5,000 crime scene samples are added in a typical month. In 2002/03, the matches reported to the police identified one or more suspects for 43,904 crime scenes. This was a 36% increase over the previous year.

The chance of obtaining one or more suspects for an offence when a crime sample is first added to the database currently stands at about 45%. There is a further probability, of about 25%, of a match being obtained later as new subject sample profiles are added to the database. The chance of identifying a crime scene with an individual when a subject sample is first added to the database is about 4%.

Familial searching
If a crime profile is checked against The National DNA Database® and no match is obtained, it is possible, because there is a tendency for criminality to run in families, that the profile of a close relative may be on the database.

The FSS has developed a technique for searching The National DNA Database® for profiles that could relate to the father/child and siblings of the offender. In combination with other information the police may have available, it is possible to reduce the results of this search to a manageable number of individuals for further investigation.

As previously mentioned, the advanced techniques of Y-STR profiling and mtDNA sequencing can be brought into play to investigate families identified by this process.

This approach is of particular value in major crime investigations. Further details are available from the FSS Forensic Intelligence Bureau (FIB), for contact numbers (see page 33).
Principles for the evaluation of DNA evidence

Assessing the value of the DNA evidence

If it is found that the DNA profile of a suspect is different from that from a crime sample, then it is reasonable (processing errors excepted) and non-controversial to conclude that the DNA in the crime sample is not that of the suspect.

If, however, the two profiles are indistinguishable - a match - then things are not nearly so straightforward. An obvious question is: Does that mean that the crime sample was left by the suspect? Unfortunately there is not a simple answer. It is most helpful to view the issues within the context of possible future proceedings. Then, if the origin of the DNA in the crime sample is in dispute, prosecution and defence will have opposing viewpoints.

We could expect the prosecution to invite the jury to believe a proposition of the following kind:

*The DNA in the crime sample came from the defendant.*

The defence view would be that the DNA in the crime sample did NOT come from the defendant. Since the DNA must have come from someone we could formulate a defence proposition of the following kind:

*The DNA in the crime sample came from some unknown person.*

The jurors will, ultimately, have to decide which of these two propositions they believe to be the correct one. They will have heard not only of the DNA match, but also of other non-scientific evidence, including evidence of opportunity (i.e. whether the defendant could have been at the scene of crime at the
appropriate time) and whatever the defendant says in his/her defence (e.g. his/her alibi). The important question is how the scientist can assist the jury in a logical, balanced and robust fashion to reach their conclusion.

It is now widely accepted among statisticians and legal scholars who have considered the problem that the scientist should, in the present context, address the following questions of the kind:

What is the probability of a DNA match if the prosecution proposition were true?

What is the probability of a DNA match if the defence proposition were true?

In this particular case (though not so in the case of DNA mixtures) the answer to the first question is quite simple - a match is exactly what we would expect to see. So we are left with the second question which we can rephrase as:

What is the probability of a DNA match if the DNA in the crime sample had come from some unknown person?

This is not such a simple question because it is necessary to clarify what is meant by an “unknown person”. As each person’s DNA is inherited from their mother and father it follows that people who are closely related are more likely to share the same profile than people who are completely unrelated. Indeed, identical twins will almost certainly have the same profile. The probability that two full siblings will have the same SGM Plus profile is of the order one in 10,000 and the corresponding probability for a pair of first cousins is of the order one in 100 million. For unrelated people, it is current FSS policy to quote a match probability of one in a billion - for reasons explained later.

It follows from this that, if the scientist is going to help the jury effectively, he/she needs to be provided with whatever information is available to sharpen up the notion of an “unknown person”. If it is known, for example, that the defendant has close siblings who
might realistically be considered as suspects for the offence, then the best way to minimise uncertainty would be to take DNA samples from those siblings to see if they can be eliminated. If that is not possible, then the scientist can guide the jury with the various relevant match probabilities. But it would be wrong to suggest that the jury have a simple task.

If we assume that the circumstances are such that close relatives are not an issue, then the question to be addressed in relation to the defence proposition is:

*What is the probability of a DNA match if the DNA in the crime sample had come from some unknown person, unrelated to the defendant?*

As the current FSS policy is to quote a match probability in answer to this question of the order one in a billion, written in full, the sort of statement the scientist would make is:

*If the DNA in the crime sample had come from some unknown person unrelated to the defendant, the probability of a match would be of the order one in a billion.*

**Match probability of one in a billion**

We have seen that a full SGM Plus profile consists of ten loci (we leave the gender marker to one side for the time being) and at each locus two alleles will be identified - one from the individual's mother and one from the father. If we wished to calculate a match probability for a single locus, we would refer to a database of samples taken from people from a relevant population to estimate how rare or how common the two components are. Then we would carry out a calculation of a match probability for that particular locus. The database that we would use need not be very large - typically one of 100 to 200 individuals is sufficient to provide robust estimates for single locus match probabilities. The calculation is carried out using formulae that are recognised as sensible by a large international consensus of scientists and statisticians. The formulae include a
correction factor that allows for the possibility that human populations may be, to a small extent, structured by the existence of groups that tend not to mix with other groups.

So, given a suitable database, we could carry out a match probability calculation for each of the ten loci - these individual match probabilities would be of the order one in 10 to one in 100.

We then face the question of how we are to combine these probabilities. There is an international consensus in favour of the practice of multiplying them all together and this leads to unbelievably small match probabilities - perhaps of the order one in a trillion (one followed by 12 zeros)) or one in a quadrillion (one followed by 15 zeros).

The FSS, however, does not follow this practice because it is considered that current understanding of the population genetic factors involved does not warrant such extravagantly infinitesimally small numbers. Instead, if we have a match between two full SGM Plus profiles we do not carry out a case specific match probability calculation. Instead, if we are considering the alternative proposition of an unknown person, unrelated to the suspect, then we quote a match probability “of the order one in a billion” (a billion is a thousand million, or one followed by nine zeros). This is independent of any particular database and we quote the same figure whatever the ethnic group of the unknown person. The justification for this procedure was published in the open literature several years ago and, to our knowledge, it has never been questioned.

Reporting a DNA match to court

As we have seen, interpreting the significance of a match between two DNA profiles depends on the alternative propositions to be addressed. If, at the time of writing a statement, the scientist is presented with a clear defence alternative, the appropriate match probability will be quoted. However, in many cases this is not the position and the question of possible relatives of the suspect is not clear to the
scientist. In such cases, the FSS will report the match without providing a single match probability. Instead, a range of match probabilities for people of different degrees of relationship with the suspect will be given.

If, later, court proceedings are envisaged and the position with regard to alternative propositions becomes clearer, then the scientist will provide an additional statement that addresses the appropriate propositions by means of the appropriate match probabilities.

Uniqueness probability

A question that is often asked is: Can a DNA match prove identity in the absence of any other evidence? But is there ever a case in which there is no other evidence? At the very least, there will be evidence relating to time/place of the crime and some information, however vague and unspecific, relating to the defendant's whereabouts around that time.

There may indeed, as in *R v Dennis John Adams*, be eyewitness and alibi evidence on behalf of the defendant. The Appeal Court, in *R v G Adams and Doheny*, took a rather simplistic view of the problem by suggesting that it was sensible to consider the number of other people in some population or other who might be expected to possess the same given profile as the defendant.

A rather naive approach would be to ask the scientist: How many other people in the world would be expected to possess the same DNA profile as the defendant? Putting the issue of close relatives to one side for the present, a match probability of one in a billion implies that one would expect there to be of the order ten other people in the world with the same profile. It is difficult to see how this helps, however, because there will never be a case in which it would be realistic to consider the entire population of the world as potential suspects for a given crime. If the crime is a rape in Huddersfield, would we seriously consider a 70-year-old female resident of Beijing a suspect for the crime? This consideration appears to have been recognised in *R v Doheny and Adams*. 
where it was suggested that a starting point might be the British population - or, in a rape case, the male population of the country. But here too, it would seem unrealistic in a rape case to consider all the males of the country as suspects.

The judgement in the above case suggests addressing a more restricted population, determined by the case circumstances. The judgement gives the sexually active male population of Manchester as an example. This brings us to another limitation. Imagine that the court has decided that it is reasonable, given the case circumstances, to consider a population of a million males. Then, in answer to the question: How many other males in this population would be expected to share the given DNA profile?, the match probability of one in a billion suggests the answer is one thousandth of a person. This clearly would be something that a jury may be expected to have difficulty in understanding and we need to follow a better approach to the problem.

It is possible to do this by calculating what is known as a “uniqueness probability”. To do this, the scientist would need to be given a guide with regard to the size and nature of the population to consider. Here it is necessary to include consideration of the possibility of close relatives. If, for example, the defendant has two brothers, who have not been excluded by profiling, yet who might be realistically be considered to be alternative suspects, then an allowance for two brothers can be included in the calculation. Also, the scientist could factor in a generous provision for relatives such as uncles and first cousins.

\[^1\] Doheny and Adams [1997] 1 Cr. App. R 369. Confusingly, the judgement adopted the phrase “random occurrence ratio” which is not only inaccurate but also has no technical or scientific pedigree. There are a few phrases that are more accurate but the consensus view among British statisticians who have studied this field is that “match probability” is the most appropriate. From a scientific viewpoint, the phrase “random occurrence ratio” should be avoided.
Scientist’s statement
In the past, it has been frequent practice for the scientist to add some sort of verbal explanation (such as “extremely strong support”) to convey the weight of evidence to be associated with a DNA match. For various reasons this practice is being discouraged. The statement that the scientist will initially provide will depend on the information available. If there is a clear pair of competing alternative propositions the scientist will quote the appropriate match probability. In a case where there is no question of close relatives of the suspect to be considered then this will be “of the order one in a billion”. If there is no clear alternative, and particularly in the case of no comment from the suspect, the match will be reported by means of an abbreviated statement together with a standard paragraph explaining the ranges of match probabilities that might be relevant.

In the event of clear competing alternative propositions later emerging, the scientist can provide an additional appropriately worded statement taking account of the updated information.

The prosecutor’s fallacy
Presenting probabilistic evidence in court has many pitfalls but there is one, in particular, that has led to successful appeals (e.g. in Doheny and Adams). Consider, for example, where there is a full SGM Plus match between the defendant and semen recovered from a rape victim and, for simplicity, assume that all considerations of possible close relatives have been excluded by one means or another. Then the scientist might say:

“The probability of a match if the semen came from another person is one in a billion.”

In an effort to simplify things for the benefit of the jury, counsel might be tempted to invite the scientist to agree to the following paraphrase:

The probability that the semen came from another person is one in a billion.
The second sentence is undoubtedly simpler but it does not follow from the first sentence. This is what is known as the “prosecutor’s fallacy”, though statisticians know it as the “fallacy of the transposed conditional”. This is, undoubtedly, what the learned judges had in mind in the judgement in *Doheny and Adams* where they say:

“The 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.”

As a general rule, the role of the scientist is to advise the court of the probability of the evidence given the proposition, in this example, of the probability of a match if the semen came from some unknown person. It is properly the role of the jury to consider the probability of the proposition given the evidence, in this example, the probability that the semen did come from some unknown person.

**Hierarchy of propositions**

Over recent years, the FSS has given much attention to forming a clear logical framework for presenting scientific evidence in a balanced and robust fashion. This is being taken forward through a project entitled *Case Assessment and Interpretation* (CAI). One of the key ideas to arise from this project work has been the formulation of a “hierarchy of propositions”. Consider the following set of propositions:

i)  *The defendant assaulted Mr Z.*  
*The defendant had nothing to do with the assault of Mr Z.*

ii)  *The defendant kicked Mr Z in the head.*  
*The defendant was not present when Mr Z was kicked in the head.*

iii)  *The blood on the defendant’s clothing came from Mr Z.*  
*The blood on the defendant’s clothing came from some unknown person.*
Here we see a progression. The first pair are clearly the level of propositions that might be addressed by a jury and are thus called **offence level** propositions. The pair of propositions at (ii) are termed **activity level** propositions. Finally the progression moves down to a level at (iii) that is comparable to those considered so far in this guide, termed **source level** propositions.

Clearly, the scientist has a role in assisting the court in explaining how the scientific evidence relates to a pair of source level propositions. But there might also be cases where, because of a combination of circumstances, the scientist might be able to assist in addressing activity level propositions. In the above case, for example, the scientist might be able to bring expertise to bear on the interpretation of the pattern of the blood staining on the defendant's clothing.

In general, the higher up the level of propositions that can be addressed, i.e. closest to offence level, the greater the assistance the scientist can give to the court. However, it is essential to realise that the level at which the scientist can assist is determined to a large extent by the information that is made available to him/her. In particular, it is important to know anything that the defendant might be saying that is relevant to formulating the propositions.

With the increase in sensitivity in DNA profiling that has been seen in recent years, scientists have had to allow for another eventuality. In certain circumstances it may not be possible for the scientist to be confident that the DNA profile that has been generated has actually come from the biological material that it is thought has been tested. For example, if this material is a degraded or weak stain and has not given a DNA profile but an underlying body fluid or a trace of cellular DNA has given a profile.

Hence we also have **sub level one** propositions:

*The DNA on the defendant's clothing came from Mr Z.*

*The DNA on the defendant's clothing came from some unknown person.*
The evaluation of DNA evidence in database matches

The overwhelming success of The National DNA Database® in the UK is now well known, not only by all those involved in the criminal justice system, but also in the public eye, due to numerous convictions in high profile cases.

All new profiles (from CJ samples, suspects, or undetected crime scene stains) loaded onto The National DNA Database® are searched against all the profiles already held and also against each other. The police are then automatically informed of any matches. The term match rather than hit is preferred, as the latter tends to imply a profile being connected to a crime(s) which may not always be the case. The information from The National DNA Database® is for intelligence purposes only, so any matches linking a suspect to a scene must be followed up by taking a control sample (usually a buccal scrape) from the suspect to confirm the result and to be used in evidence.

Given the size of the The National DNA Database® it is quite possible that SGM profiles from two unrelated individuals will match by chance, this is a so-called ‘adventitious match’. However, given the greatly increased discriminating power, such occurrences can be expected to be very rare with SGM Plus.

Some press reports have suggested that these adventitious matches are mistakes that may have led to the conviction of innocent people. They are not mistakes, they are predictable occurrences. The mistake occurs if the possibility of the match being adventitious is not considered and individuals are arrested and prosecuted without first investigating alibis, motives, geographical opportunity or, where appropriate, upgrading the profiles involved to SGM Plus profiles to reduce the risk of the match being adventitious.

Police must therefore be aware of the consequences of adventitious matches and that DNA evidence is not used in criminal prosecution without careful
consideration of the surrounding circumstances. Indeed, this was one of the findings in the Appeal Court hearing of *R. v Lashley* (2000), where DNA evidence had been relied upon for a conviction. There was no evidence that the appellant had been in the vicinity of where the crime had been committed and hence the DNA (SGM only) profile was just as likely to have been left by any one of five or six other men in the United Kingdom.
Presentation of the evidence and statements

DNA evidence has been presented successfully in courts since 1988. Over that time the techniques involved in DNA profiling have been continuously changed and improved.

The primary means of communicating the findings of the forensic scientist to the court is through statements. The format and style in which the statement is written is vitally important and these are also kept constantly under review by the FSS. The introduction of new techniques as well as developments in the criminal justice system may well require changes to be made in the future.

Areas of challenge initially concentrated on the continuity of sample handling, the reliability of the actual techniques and the strength of DNA match probabilities. Now, however, they are more to do with the relevance of the DNA evidence in relation to the rest of the case when taken into account with the other, non-DNA, evidence.

A change to Crown Prosecution Service (CPS) policy that provides for suspects to be charged on the basis of an intelligence match on The National DNA Database®, providing there is some further supporting evidence, has been introduced (CPS Policy Directorate, August 2004). The implementation of this policy is supported by guidance provided by both the CPS and the Association of Chief Police Officers. Where this prosecution process is being followed further evidential material will only be requested from a forensic scientist where “no indication of plea” or a “not guilty plea” has been entered. Such additional work will be performed in stages at the request of the prosecution team. On submission of the reference sample an abbreviated statement confirming the match will be provided. Further statements dealing with continuity and hearsay can be requested where admissions on these matters are not obtained from the defence. Full
evaluative statements will normally only be provided where the DNA evidence is to be contested at trial and after a defence statement is available.

The FSS recommends that if it becomes apparent that the DNA evidence will be a significant issue at trial the scientist is contacted. We believe that it is vital that the scientist knows the background information, and any other relevant evidence, in order that appropriate alternatives can be considered. Case conferences are an ideal way of talking through the issues, setting the DNA evidence into context, deciding if further work or analysis is required and discussing how the evidence could be presented. If visual aids are required for the presentation of the evidence, the form of these and the resources required can also be discussed at the case conference.

Further advice and suggested reading

Further advice on any particular case can be obtained from the scientist concerned.

General enquiries and requests for further copies of this booklet can be directed to the FSS customer services helpdesk on 0121 607 6996/7/8.

Specific enquiries relating to The National DNA Database®, e.g. match reports, sample retention policies and submissions should be directed to The National DNA Database® helpdesk on 0121 606 2950.

Other booklets available from the FSS, which supplement this booklet, include:

- **DNA Present and Correct - A guide to DNA Profiling**
- **The Scenes of Crime Handbook 2004**
- **DNA Profiling: A discussion of issues relating to the reporting of very small match probabilities - Criminal Law Review; 2000, 341-355**
- **Lawyers’ Guide to DNA (previous version of this booklet)**
Appendix

Glossary of terms and abbreviations

Adventitious match DNA profiles from two individuals, who are not identical twins, that match by chance.

CJ (criminal justice) sample Non-intimate sample (cheek scrape or hair root) lawfully taken from a suspect arrested, charged or convicted for a recordable offence for intelligence purposes.

Reference sample Sample taken from a person for comparison with a specific crime sample and used as evidence in a court of law.

Crime sample Sample recovered from a scene of crime or victim from which DNA can be obtained for intelligence or evidential use.

DNA Deoxyribonucleic Acid.

Familial searching Utilises the fact that elements of DNA are inherited. Unknown crime profiles can be searched against The National DNA Database® to return profiles of potential relatives of the offender.

FSS* The Forensic Science Service.

LCN Low Copy Number - analysis and interpretation process used for samples containing extremely small amounts of DNA.

Likelihood ratio The probability of the evidence given the prosecution proposition divided by the probability of the evidence given the defence proposition.

Match probability The probability that an unknown individual will have a particular profile, given that a known individual has that profile.

MtDNA Mitochondrial DNA.
The National DNA Database®  
Relational database holding DNA profiles from CJ samples, reference samples, volunteer samples and undetected crime samples that are searched against each other and all new profiles that are loaded onto the database.

Pendulum list search (PLS)  
A technique for identifying and searching individual DNA profiles that are contained in a two-component mixed profile.

PCR  
Polymerase chain reaction - technique used for targeting and amplifying specific regions of DNA.

SGM  
Second generation multiplex - profiling system developed and utilised by the FSS from 1995 - 1999.

SGM Plus**  
Profiling system utilised by the FSS since July 1999. A more sensitive and discriminating system than SGM but compatible with SGM.

STR  
Short tandem repeat - non-coding region of DNA targeted in the SGM/SGM Plus profiling methods which varies in length between individuals.

Y-STRs  
Short tandem repeats contained in the male Y chromosome in nuclear DNA.

* The Forensic Science Service®, FSS® and The National DNA Database® are UK registered trademarks.

**AmpFISTR® SGM Plus is a registered trademark of Applied Biosystems, a division of the Perkin Elmer Corporation.