Summer Institute in Statistical Genetics
University of Washington, Seattle

Forensic Genetics Module

# Evaluating DNA evidence

David Balding
Professor of Statistical Genetics
University of Melbourne and UCL Genetics Institute, London

July 20 and 21, 2016

## Contents

**1 Introduction to weight-of-evidence theory** **3**
1.1 Warm-up examples . . . 3
1.1.1 People *v.* Collins (California, 1968) . . . 3
1.1.2 Disease testing (Positive Predictive Value, PPV) . . . 4
1.1.3 Coloured taxis . . . 5
1.2 Rare trait identification evidence . . . 6
1.2.1 The "island" problem . . . 6
1.2.2 A first lesson from the island problem . . . 8
1.3 Making the island problem more realistic . . . 8
1.3.1 The effect of uncertainty about $p$ . . . 9
1.3.2 The effect of possible typing errors . . . 10
1.3.3 The effect of searches . . . 10
1.3.4 The effect of other evidence . . . 12
1.3.5 The effects of relatives and population subdivision . . . 13

**2 Assessing evidence using likelihood ratios** **14**
2.1 The weight-of-evidence formula . . . 14
2.1.1 Application to the island problem . . . 16
2.1.2 Two pieces of evidence . . . 16
2.1.3 Application of the formula . . . 16
2.2 Consequences for DNA evidence . . . 18
2.2.1 Many possible culprits . . . 18
2.2.2 Incorporating the non-DNA evidence . . . 18
2.2.3 Relatives . . . 19
2.2.4 Laboratory and handling errors . . . 20
2.3 Database searches . . . 21
2.4 Derivation of the weight-of-evidence formula . . . 22
2.4.1 Bayes Theorem . . . 22
2.4.2 Typing errors . . . 23

**3 Other approaches to weight of evidence** **24**
3.1 Uniqueness . . . 24
3.2 Random Man Not Excluded (RMNE) probabilities . . . 25
3.3 Hypothesis Testing . . . 26

**4 Calculating LRs allowing for coancestry** **27**
4.1 Some population genetics . . . 27
4.1.1 Population genotype probabilities . . . 27
4.1.2 A sampling formula for alleles . . . 28
4.2 Identification: single contributor . . . 30
4.2.1 Single-locus match probabilities . . . 32
4.2.2 Multiple loci: the "product rule" . . . 33
4.2.3 Relatives of $Q$ . . . 33
4.3 Issues for casework . . . 35
4.3.1 Values for the $p_j$ . . . 35
4.3.2 The value of $F_{ST}$ . . . 36
4.3.3 The hypotheses . . . 36

**5 More complex scenarios** **37**
5.1 Paternity and relatedness . . . 37
5.1.1 Single locus: paternal allele known . . . 37
5.1.2 Single locus: paternal allele unknown . . . 38
5.2 Identification: mixed profiles . . . 39
5.2.1 Visual interpretation of mixed profiles . . . 39
5.2.2 Likelihood ratios under qualitative interpretation . . . 41
5.2.3 The number of unknown contributors . . . 41
5.2.4 Two contributors: Q/X and K . . . 42

5.2.5 Two contributors: Q/X and U . . . . . 44
5.2.6 Quantitative interpretation of mixtures . . . . . 46
5.3 LRs for low template DNA profiles . . . . . 47

**6 Some legal issues** **47**
6.1 The role of the expert witness . . . . . 47
6.2 The prosecutor's fallacy . . . . . 48
6.3 The defendant's fallacy . . . . . 48

# 1 Introduction to weight-of-evidence theory

## 1.1 Warm-up examples

### 1.1.1 People *v.* Collins (California, 1968)

An apparently reliable witness testified that the crime was committed by two individuals with the characteristics given in the table below. A couple fitting the description was charged with the offence. An instructor in mathematics was called as expert witness and suggested the probabilities shown.

| Trait | Probability |
|---|---|
| Yellow car | 1/10 |
| Man with moustache | 1/4 |
| Woman with pony tail | 1/10 |
| Woman with blond hair | 1/3 |
| Black man with beard | 1/10 |
| Inter-racial couple | 1/1000 |

The mathematics instructor multiplied these probabilities together and obtained a very small number. The defendants were found guilty.

The conviction was overturned by the California Supreme Court, but it made errors in its analysis.

Many criticisms can be raised about the prosecution evidence in *Collins* and many pages of academic literature have been devoted to raising them. Much of this literature is flawed in crucial respects. We won't analyse this in any detail here, but the example raises fundamental issues such as:

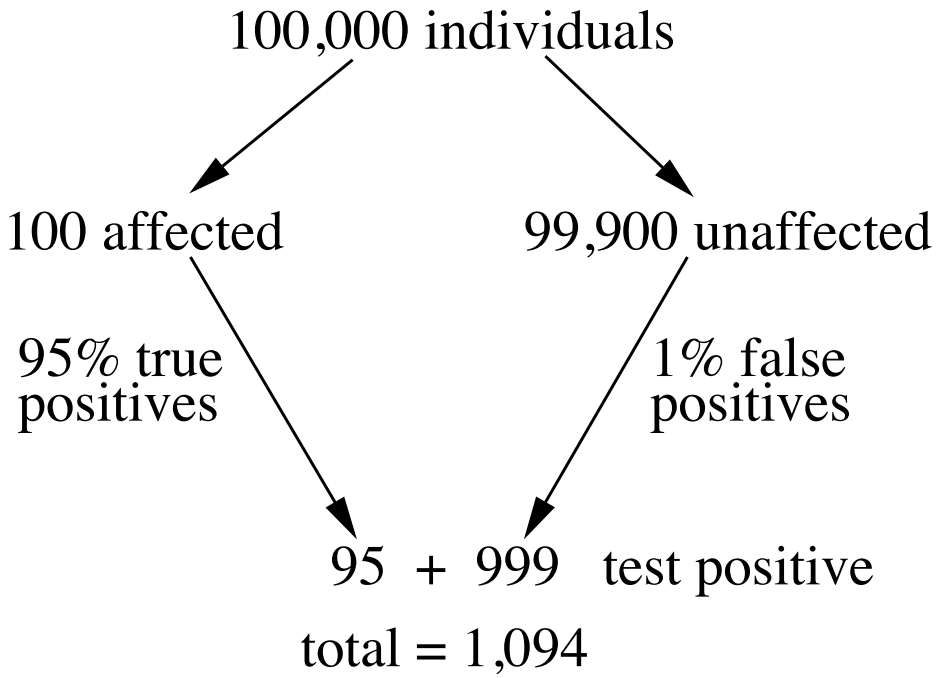
- What numbers should be presented to a court to help jurors evaluate evidential weight?
- How should they be interpreted?
- What possible errors should jurors be warned against?

### 1.1.2 Disease testing (Positive Predictive Value, PPV)

Suppose that, although showing no symptoms, you decide to take a diagnostic test for a rare disease that can exist in a latent state before showing symptoms. Let's assume the following facts:

- about 1 person in 1 000 has the latent form of the disease;
- the false positive rate is 0.01 (this is the proportion of positive outcomes among those taking the test who do not have the disease);
- the false negative rate is 0.05 (this is the proportion of negative outcomes among those taking the test who do have the disease);

The test result comes back positive. How worried should you be? What is the probability that you have the disease?




The rare disease problem is closely connected with the weight-of-evidence problem: there are two possible "states of nature", disease and no disease (compare with guilty and not guilty), and there is a diagnostic test that is reliable but can occasionally fail (cf. a DNA profile test, that can occasionally result in a match by "chance", or by a laboratory or handling error). The correct method of reasoning for rare diseases leads to results that at first are counter-intuitive for many people: it can do more harm than good to screen the general population for a rare disease, even when an accurate test is available.

The logic of the probability analysis is compelling, and its implications are now universally accepted for public health policy. The analogous reasoning for DNA profile evidence

also leads to some surprising conclusions, and after a struggle during the 1990s, they are slowly becoming accepted in the courts.

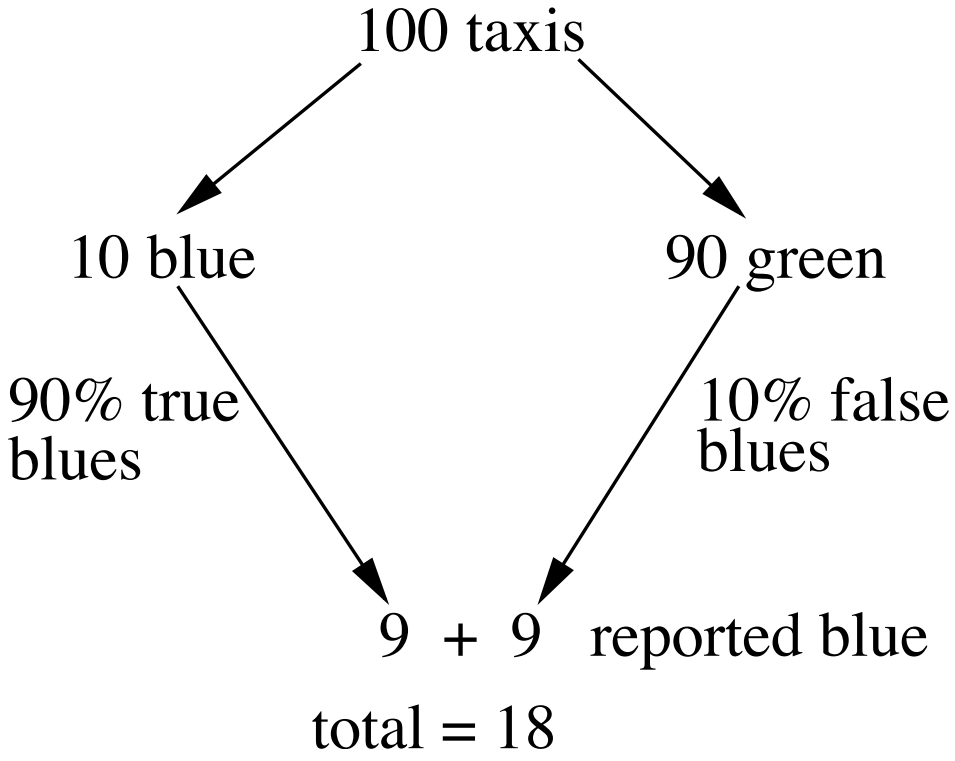
In public health, the probability of having the disease given a positive test result is called the "positive predictive value" (PPV) of the test. It is a difficult quantity because it depends on the disease prevalence – and this will vary, for example according to ethnic group and occupation. In contrast, the false positive and false negative rates are easier to work with because they can be measured in the laboratory – so many focus on these even though they do not answer the relevant question, for which the PPV is needed. Similarly in forensic identification, it is not the "match probability" for the DNA profile test that ultimately matters, but the equivalent of the PPV, and this is a difficult quantity because it depends, for example, on the other evidence in the case.

Before tackling these problems, let's consider one more example and to see if you have got the hang of these problems.

### 1.1.3 Coloured taxis

Suppose that 90% of the taxis in the town are green and the rest are blue. According to an eyewitness, the perpetrator of a "hit-and-run" traffic offence was driving a blue taxi. We assume that the eyewitness testifies honestly, but may have made a mistake about the colour of the taxi: it was dark at the time, and tests indicate that eyewitnesses mistake blue taxis for green, and vice versa, about 1 time in 10 under these conditions.

What is the probability that the taxi really was blue?



## 1.2 Rare trait identification evidence

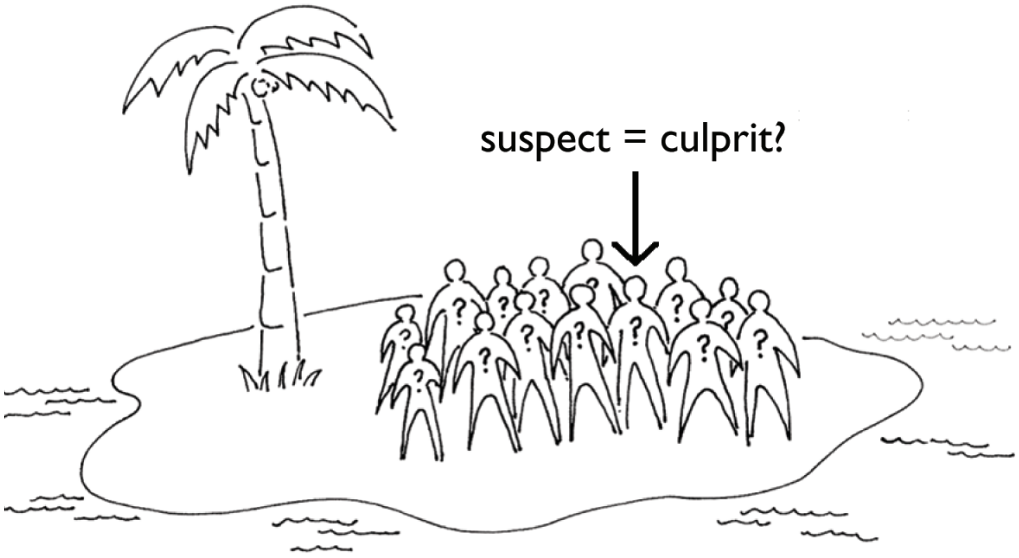
One of the methods scientists use to try to understand complex phenomena is to investigate simple models. Seeing what happens when simple models are tweaked can suggest insights into how the real world works. The models can be elaborated to bring them a little closer to reality, but it is often the simplest models that give the most profound insights.

We will take this approach to assessing the weight of DNA evidence. An actual case involves many complications:

- how many previous suspects were typed and excluded?
- what were the possibilities for a contamination error?
- are any of $Q$'s close relatives possible culprits?

and many more. We can't cope with all these complications at once. Instead, we will start with an imaginary crime on an imaginary island where life, and crime, is much simpler than in our world. Although unrealistic, analysis of the "island problem" leads to profound insights. We will gradually add more features to bring the island closer to the real world. In so doing we will learn new lessons about evidential weight.

### 1.2.1 The "island" problem




Consider a rare, latent trait: it could be a DNA profile but there is no need to be specific at this stage. Let's use the symbol $\Upsilon$ to denote the trait. A crime is committed on a remote island with a population of, say, 101. At first, there are no clues, and everyone on the island is equally under suspicion. Then it is learned that the culprit possesses $\Upsilon$,

and a suspect is identified who is found to have $\Upsilon$. How convinced should we be that the suspect is guilty?

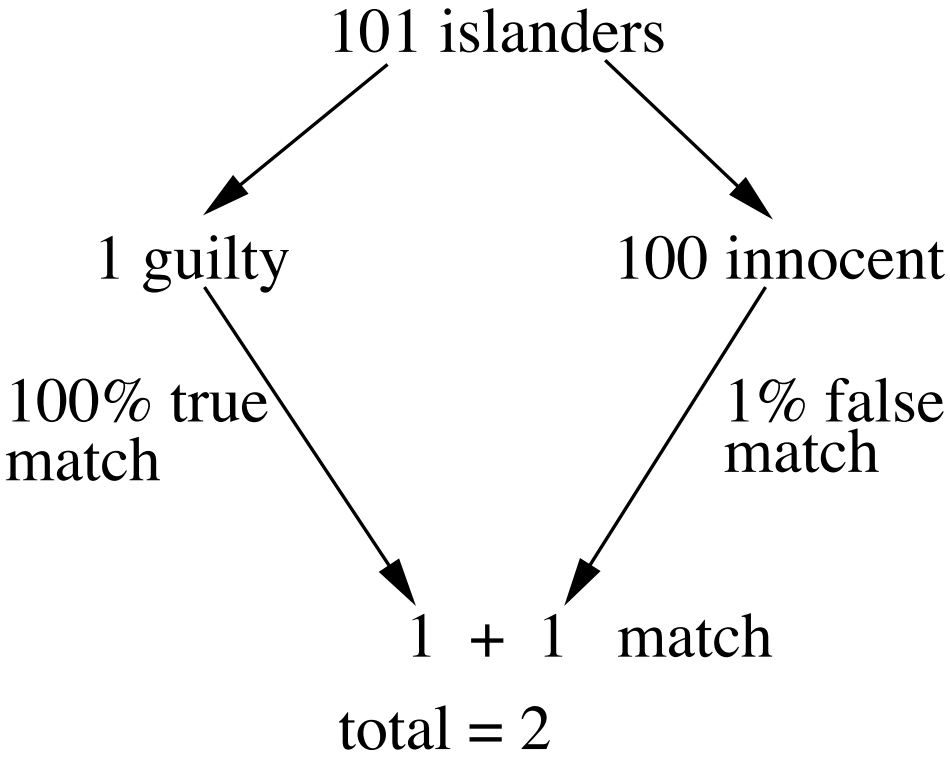
The answer to this question depends on, among other factors, how rare $\Upsilon$ is. Suppose that the suspect and the culprit (who may or may not be the same person) are the only people on the island whose $\Upsilon$-status is known. A recent survey on the nearest continent, however, indicated that 1 person in 100 has $\Upsilon$ and we assume that people on the neighbouring islands have $\Upsilon$ independently, with probability 0.01.

## The island problem: facts summary

- All 101 islanders are initially equally under suspicion;
- The culprit has $\Upsilon$;
- The suspect has $\Upsilon$;
- The $\Upsilon$-states of the other islanders are unknown;
- We expect on average about 1 person in 100 to have $\Upsilon$.

What is the probability that the suspect is guilty?

## The island problem: solution



In addition to the one individual known to match (the suspect), we expect one other individual on the island to match. In the absence of any further evidence pointing to, or away from, the suspect, the probability that the suspect is guilty is 1/2.

### 1.2.2 A first lesson from the island problem

It is the probability that the defendant is guilty given a match of $\Upsilon$-states – the analogue of the PPV for disease testing – that is directly relevant to the juror's decision. Like the PPV, this probability is, in reality, a slippery quantity to work with, but in the simplified island setting we can calculate the probability of guilt. Here, we are given that there is no other evidence and that all islanders are initially equally under suspicion, so we are permitted to ignore factors like age, state-of-health, and distance from the crime scene that complicate real-world crimes.

In general, if there are $N$ people on the island other than the suspect $Q$, and the probability that any one of them has $\Upsilon$ is $p$, then the formula for the probability that $Q$ is guilty, given the $\Upsilon$ evidence, is

$$\mathrm{P}(G|E) \equiv \mathrm{P}(Q \text{ guilty given evidence}) = \frac{1}{1 + N \times p}. \tag{1}$$

We will see how to derive this equation in Section 2.4 below. If $N = 100$ and $p = 1/100$ then

$$\mathrm{P}(G|E) = \frac{1}{1 + 100 \times 1/100} = \frac{1}{2},$$

Although it follows from (1) that the rarer is $\Upsilon$, the higher is the probability that $Q$ is guilty, the strength of the overall case against $Q$ depends on *both* the rarity of $\Upsilon$ (i.e. on $p$) and on the number of alternative possible culprits $N$.

**Lesson 1** *The fact that $\Upsilon$ is rare (i.e. $p$ is small) does not, taken alone, imply that $Q$ is likely to be guilty.*

The unrealistic aspects of the island problem are immediately apparent. Nevertheless, Lesson 1 is "robust": when we change the problem to make it more realistic, Lesson 1 still applies (see Section 2.2.1).

## 1.3 Making the island problem more realistic

Let's change some aspects of the island problem to investigate how different factors affect weight of evidence. To avoid getting too bogged down in complications, we will investigate different factors one at a time.

### 1.3.1 The effect of uncertainty about $p$

In practice, we will not know $p$ exactly. Let's modify the island problem by attaching some uncertainty to $p$, measured by a variance $\sigma^2$. The island problem formula (1) now becomes

$$\mathrm{P}(G|E) = \frac{1}{1 + N \times (p + \sigma^2/p)}, \tag{2}$$

which is always less than (1). We therefore have immediately the following important conclusion, that, like Lesson 1, turns out to be robust.

**Lesson 2** *Uncertainty about p does not "cancel out". Ignoring uncertainty is unfavourable to defendants.*

*Numerical illustration*

In the original island problem of Section 1.2.1, $N = 100$ and $p = 0{\cdot}01$, and the probability of guilt is

$$\mathrm{P}(G|E) = \frac{1}{1 + N \times p} = \frac{1}{1 + 100 \times 0{\cdot}01} = 50\%.$$

If, now, we suppose that there is some uncertainty about $p$, say $p = 0{\cdot}01 \pm 0{\cdot}005$ (i.e. $\sigma = 0{\cdot}005$), then the modified formula for the probability of guilt is

$$\begin{aligned} \mathrm{P}(G|E) &= \frac{1}{1 + N \times (p + \sigma^2/p)} \\ &= \frac{1}{1 + 100 \times (0{\cdot}01 + (0{\cdot}005)^2/0{\cdot}01)} \\ &= \frac{1}{1 + 1 + 0{\cdot}25} = \frac{4}{9} \approx 44\%. \end{aligned}$$

So failing to acknowledge the uncertainty about $p$ overstates the probability of guilt; here, 50% instead of the correct 44%.

Uncertainty about $p$ can arise for a number of reasons. If knowledge about $p$ comes from a survey, then there is uncertainty due to sampling: a different sample would have led to somewhat different results. Moreover, there is always uncertainty due to the possibility that the sample is unrepresentative (the islanders may differ from the population of the continent).

We will see in Section 4 that uncertainty about $p$ is crucial to the correct interpretation of DNA evidence: uncertainty enters because of sampling error but also because DNA profile frequencies vary among ethnic/religious/social groups, and we never know exactly which is the correct reference group in a particular case, or what are the profile frequencies

in the relevant groups. The rarity of a DNA profile is assessed in part on the basis of data, from "convenience" samples rather than scientific random samples, and in part on the basis of population genetics theory, that holds at best only approximately in actual human populations. In practice, then, the $\sigma^2/p$ in (2) is often much larger than the $p$, so that ignoring the former can be much more important than in the illustration above.

### 1.3.2 The effect of possible typing errors

Now forget uncertainty, and assume again that we know $p$ exactly. Suppose, however, that there is a probability $\epsilon_1$ that an individual who does not have $\Upsilon$ will be wrongly recorded as having it (i.e. a false positive), and probability $\epsilon_2$ for the other, false negative, error. We assume that these probabilities apply for typing both $Q$ and the culprit, and that errors occur independently.

These assumptions are still unrealistically simple, but they allow a first insight into how the possibility of error affects evidential weight. The exact formula is a little complex, but an approximation appropriate here is given by

$$\mathrm{P}(G|E) = \frac{1}{1 + N \times (p + \epsilon_1)^2/p}. \tag{3}$$

We will discuss the derivation of (3) in Section 2.4.2 below. This simple formula suffices for another important, robust, lesson.

**Lesson 3** *The overall weight of evidence against $Q$ involves adding together the probability of a "chance match" and the probability of a match due to a typing error.*

Laboratory and handling errors for DNA profile evidence are discussed further in Section 2.2.4. In general, the value of $\epsilon_1$ is difficult to assess, usually more difficult than $p$ since it depends strongly on the circumstances of a particular case. Ultimately, it is for the jury in criminal trials to assess the probability that some error has occurred, on the basis of the evidence presented to it. It is important that courts be given some idea of what errors are possible, how likely they are in the present case, and what effect possible errors have on evidential weight.

### 1.3.3 The effect of searches

Forget, for the moment, uncertainty and errors, and focus on a new issue. In the island problem, the question of how $Q$ came to the attention of the crime investigators was ignored. This is not as unrealistic as it may first appear: in practice, suspects are often identified on the basis of a combination of factors such as previous convictions, suspicious

behaviour, criminal associates and so forth. Such reasons may not form part of the evidence presented in court, in which case, as far as a juror is concerned, $Q$ "just happened" to come to the attention of the authorities. Further, the legal maxim "innocent until proven guilty" is usually interpreted to mean that, before the evidence is presented in court, $Q$ should be regarded as being just as likely to be guilty as anybody else in the population.

Suppose now that $Q$ was identified on the basis of a search for $\Upsilon$-bearers. That is, the islanders are examined in random order until the first one is found who has $\Upsilon$. This individual is then accused of the crime. As well as the facts listed in the summary on page 7, we now have the additional information that, say, $k$ islanders have been investigated and found not to have $\Upsilon$. In the original island problem, the reasons for first identifying $Q$ were not based on $\Upsilon$-possession.

**Is $\mathrm{P}(G|E)$, the probability that $Q$ is guilty, higher or lower following a search, compared with the original setting?**

There seem to be two reasons for believing that it should be lower:

1. the fact that $Q$ was initially just one person in a random sequence of individuals searched means that he/she is less likely to be the culprit;

2. if you set out to find a suspect who has $\Upsilon$, then the fact that $Q$ has $\Upsilon$ is unsurprising and therefore of little or no evidential value.

It turns out that the probability that $Q$ is guilty is *higher* following a search than in the original island problem. In fact,

$$\mathrm{P}(G|E) = \frac{1}{1 + (N-k) \times p}, \tag{4}$$

which is greater than (1). Many people, particularly scientists, find this result counter-intuitive, perhaps because of the two "reasons" given on page 11.

- The first "reason" is easily dismissed: "innocent until proven guilty" implies that every defendant should be treated, before the evidence is presented, as just another member of the population. This view is incorporated into the island problem by initially regarding every islander as equally under suspicion. So there is nothing special about a suspect identified on the basis of a search.

- To see that the second "reason" is wrong, think about the case that $Q$ was the last person searched: everyone else was inspected and found not to have $\Upsilon$. Then the fact that $Q$ has $\Upsilon$ is overwhelmingly strong evidence against him/her, as is reflected by the value $\mathrm{P}(G|E) = 1$ that is obtained in (4) when $N = k$. The key is to keep attention fixed on the relevant question, which is not "how likely is it that I will find a $\Upsilon$-bearer if I look for one?", but "given that a $\Upsilon$-bearer has been found, how strong is the evidence against this individual?".

The reason behind the correct formula is that each individual found not to have $\Upsilon$ is excluded from suspicion (remember that we are ignoring error here). The removal of this individual leaves a smaller pool of possible culprits and hence each remaining person in the pool becomes (slightly) more likely to be guilty. Notice that if the first person is found to have $\Upsilon$, so that $k = 0$, then the original island problem formula is recovered. The fact that a search was intended then makes no difference to the strength of the evidence.

**Lesson 4** *In the case of a search of possible culprits to find a "match" with crime scene evidence, the longer the search (i.e. the more individuals found not to match) the stronger the case against the one who is found to match.*

The important, related issue of suspects identified through searches of DNA profile databases is discussed below in Section 2.3.

### 1.3.4 The effect of other evidence

In the island problem, we assumed in effect that there was no evidence other than the $\Upsilon$-evidence. In practice, of course, even if there is no further evidence that is directly incriminating, there will be background information presented to the jury, such as the location, time and nature of the crime, that makes some individuals more plausible culprits than others.

**Definition 1** *We write $w_X$ for the weight of the non-$\Upsilon$ evidence against person $X$,* relative to its weight against $Q$.

In the original island problem, each $w_X$ was equal to one. A value $w_X > 1$ indicates that, ignoring the $\Upsilon$-evidence, individual $X$ is more likely to be the culprit than is $Q$. As an example, suppose that, other than the information about $\Upsilon$, the evidence consists of the location of the crime and the locations of the homes of all the islanders. A juror may reason that individuals who live near to the crime scene are more likely to be the culprit than, say, individuals who live on the other side of the island. Such an assessment can be reflected by values of $w_X$ greater than one for those who live nearer to the crime scene than $Q$, and less than one for those who live further away.

When other evidence is taken into account, the island problem formula (1) becomes

$$\mathrm{P}(G|E) = \frac{1}{1 + p \times \sum_{X=1}^{N} w_X}, \tag{5}$$

in which we introduce the mathematical symbol $\sum$ to denote summation. If all the $w_X$ are equal to one then $\sum_{X=1}^{N} w_X = N$, and the formula reduces to the original island problem formula.

The role of the $w_X$ in connection with assessing DNA evidence is discussed further below in Section 2.2.2.

### 1.3.5 The effects of relatives and population subdivision

Even though we are initially ignorant about who on the island has $\Upsilon$, the observation that $Q$ has it can be informative about whether or not other individuals, such as relatives and associates, also have $\Upsilon$. For example, if the population of the island is divided into "easties" and "westies", then the fact that $Q$, an eastie, has $\Upsilon$ may make it more likely that other easties also have $\Upsilon$.

In the island problem, we assumed that $\Upsilon$ possession for different individuals was independent, so that one person's $\Upsilon$-status carries no information about the $\Upsilon$-states of other individuals. In practice, however, this "learning" effect can be important, particularly for DNA profile evidence. Any particular DNA profile is very rare, but once that profile is observed, it becomes much more likely that other people, among the individual's relatives or ethnic group, also have it. It follows that what matters in practical cases involving DNA evidence is not $p$, the overall frequency of the profile, but the probabilities of the other possible culprits having the profile *given that* $Q$ has it.

**Definition 2** *We write $r_X$ for the* match probability *for possible culprit $X$, which is the probability that $X$ has $\Upsilon$,* given that *$Q$ has $\Upsilon$.*

Some writers on DNA evidence use "match probability" to denote the relative frequency of $\Upsilon$ in some population. This is incorrect because the concept of "match" involves two individuals, not one. In the original island problem we assumed independence of $\Upsilon$-states so that $r_X$ does equal $p$ on the island. For DNA profile evidence in real populations, however, relatedness and population subdivision mean that the match probability $r_X$ exceeds the profile relative frequency $p$, often substantially. Thus confusing $r_X$ with $p$ is detrimental to defendants (see Section 4).

When these effects are taken into account, the island problem formula (1) becomes

$$\mathrm{P}(G|E) = \frac{1}{1 + \sum_{X=1}^{N} r_X}. \tag{6}$$

If all the $r_X$ are equal to $p$, then $\sum_{X=1}^{N} r_X = Np$, and (1) is recovered.

# 2 Assessing evidence using likelihood ratios

There are many other factors that could be introduced into the island problem in order to investigate their effect on evidential strength:

- what if the culprit is not, after all, the source of the DNA obtained from the crime scene?
- what about the fact that $Q$ failed to produce a convincing alibi?
- what if the police accuse $Q$ of every crime that occurs on the island?
- what if only a few individuals could have visited the crime scene during the time of the offence?

We will now introduce a general formula for quantitatively assessing evidence in the light of such factors.

## 2.1 The weight-of-evidence formula

Although we have given some intuitive explanation of (1) through (6), we have not yet explained how to derive such formulas. The match probability $r_X$, defined on page 13, is a special case of a *likelihood ratio* (LR).

**Definition 3** *Let $E_d$ stand for some evidence and $X$ for the name of a possible perpetrator of the crime (other than the defendant, $Q$). Let $H_X$ denote the hypothesis that $X$ was the perpetrator. The* likelihood ratio *for comparing $H_X$ with $H_Q$ on the basis of evidence $E_d$ is the ratio of how likely it is to have observed $E_d$ under $H_X$ to how likely $E_d$ is under $H_Q$:*

$$R_X = \frac{\mathrm{P}(E_d|H_X)}{\mathrm{P}(E_d|H_Q)}. \tag{7}$$

We have introduced here the notation "|", which is mathematical shorthand for "given that". Most authors define the LR the other way around (i.e. with the top and bottom lines interchanged). Either definition is acceptable given the obvious adjustments to formulas. For us, $R_X$ is small when $E_d$ provides strong evidence in favour of guilt ($H_Q$). Our definition has the advantage that $R_X$ can often be interpreted as a conditional probability (see Section 4).

**Definition 4** *The* other evidence *ratio $w_X$, introduced informally on page 12, is the probability of $H_X$ divided by the probability of $H_Q$, both evaluated in the light of $E_o$, all the evidence other than $E_d$. That is,*

$$w_X(E_o) = \frac{\mathrm{P}(H_X|E_o)}{\mathrm{P}(H_Q|E_o)}.$$

Putting together the factors discussed in Sections 1.3.4 and 1.3.5, with the more general $R_X$ replacing the $r_X$, we obtain:

**The weight-of-evidence formula**

$$\mathrm{P}(G|E_d, E_o) = \frac{1}{1 + \sum_{X \in \mathcal{P}} w_X R_X}, \tag{8}$$

Formula (8) is a special case of a result in probability theory known as *Bayes Theorem*, in honour of the c18 clergyman Thomas Bayes.

*The population $\mathcal{P}$*

The summation in (8) is over some population $\mathcal{P}$ of unprofiled individuals, assumed to include all the possible sources of the crime stain other than $Q$. Although $\mathcal{P}$ should include all realistic alternative suspects, there is some flexibility as to how many extra individuals are included. Often it might be appropriate to include in $\mathcal{P}$ all individuals aged, say, between 16 and 65 living within, say, one hour driving time of the crime scene. Alternatively, $\mathcal{P}$ might include all adult male residents of the nation in which the crime occurred. However, $\mathcal{P}$ could include everyone on earth except $Q$, if desired: for a crime committed in Marrakesh, the value of $w_X$ will be very close to zero when $X$ is a resident of Pyongyang. This individual can be included in $\mathcal{P}$, but the error resulting from simply ignoring all the residents of Pyongyang will usually be negligible.

*Grouping the $R_X$*

Although there is, in principle, a separate $R_X$ for *every* person not excluded from being a possible culprit, in practice there will be large groups of individuals for whom the evidence $E_d$ bears the same weight, and thus for whom $R_X$ will take the same value. By grouping together members of $\mathcal{P}$ having approximately the same value of $R_X$, it will typically be satisfactory in practice to consider just a few distinct terms in the summation of (8).

### 2.1.1 Application to the island problem

In the island problem setting, the evidence $E_d$ can be summarised by

$$E_d = \text{“both } Q \text{ (suspect/defendant) and culprit are observed to have } \Upsilon\text{”}.$$

Since typing error is assumed impossible, $E_d$ implies that suspect and culprit do both have $\Upsilon$, in which case each LR $R_X$ is equivalent to the match probability $r_X$. The principle distinction between the two is that $R_X$ in principle allows the possibility for the evidence to have arisen by other means, such as handling error or fraud.

If $Q$ is not guilty, then we have observed two $\Upsilon$-bearers on the island: the culprit and $Q$. Under the assumptions introduced on page 8, the probability that any two individuals both have $\Upsilon$ is $p \times p = p^2$. On the other hand, if $Q$ is guilty, then we have observed only one $\Upsilon$-bearer and this observation has probability $p$. The LR for any possible culprit $X$ is thus

$$R_X = \frac{p^2}{p} = p. \tag{9}$$

Substituting this value into (8) we recover (5), and in the case that all the $w_X$ are equal to one, we once again obtain the original island problem formula (1).

### 2.1.2 Two pieces of evidence

When the evidence to be assessed, $E_d$, consists of two items, say $E_1$ and $E_2$, the LR can be calculated in two equivalent ways, corresponding to the two possible orderings of $E_1$ and $E_2$, but these give the same result. Thus, applying the weight-of-evidence formula to $E_1$ and $E_2$ together, given background information $E_o$, gives the same result as applying it to $E_2$ when $E_1$ is included with $E_o$ or to $E_1$ when $E_2$ is included with $E_o$.

Apparently strong evidence may be of little value if it merely replicates previous evidence. If $E_1$ and $E_2$ are highly correlated (e.g. matches at tightly linked genetic loci, or statements from two friends who witnessed the crime together and discussed it afterwards), then $R_X(E_1|E_2, E_o)$ and $R_X(E_2|E_1, E_o)$ may both be close to one (i.e. little evidential weight) even though both $R_X(E_1|E_o)$ and $R_X(E_2|E_o)$ indicate strong evidence. In this case the joint weight of the two pieces of evidence is about the same as the weight of either piece of evidence taken alone. On the other hand, if the items of evidence are independent given $E_o$ then

$$R_X(E_1, E_2|E_o) = R_X(E_1|E_o)R_X(E_2|E_o).$$

### 2.1.3 Application of the formula

The weight-of-evidence formula (8) can be used to assess all the evidence in a case. Suppose that the evidence can be allocated into four categories:

1. $E_1$ is information about the nature and location of the crime;

2. $E_2$ is an eyewitness description of the crime;

3. $E_3$ is the defendant's testimony;

4. $E_d$ is the DNA evidence.

Initially, the jurors might assign values to the $w_X$ based only on $E_1$. After hearing $E_2$, the juror can calculate $R_X(E_2|E_1)w_X(E_1)$. The probability of guilt $\mathrm{P}(G|E_2, E_1)$ can be evaluated at this point, if desired. If $E_3$ is now taken into account, a new probability of guilt $\mathrm{P}(G|E_3, E_2, E_1)$ can be calculated based on the values of $R_X(E_3|E_2, E_1)R_X(E_2|E_1)w_X(E_1)$ for all $X$. As noted above, these values can be regarded as $w_X(E_o)$ for the purposes of assessing the DNA evidence $E_d$, where $E_o$ stands for all the non-DNA evidence, $E_1$, $E_2$, and $E_3$.

This logical analysis of evidence is particularly useful when some items of evidence are strongly incriminating while others are exculpatory. The fact that the order in which different items of evidence is assessed does not affect the final answer is crucial. Moreover, how the evidence is categorised into "items" is also irrelevant: since

$$R_X(E_3, E_2|E_1) = R_X(E_3|E_2, E_1)R_X(E_2|E_1),$$

so the same answer is obtained whether $E_3$ and $E_2$ are analysed together, or separately.

We will focus on applications of the formula to DNA evidence. We will assume that $E_d$ refers to the DNA evidence, and that $E_o$ includes all other evidence, so that we regard the DNA evidence as being assessed last. This is for convenience and is not necessary; assessing the DNA evidence first has some advantages, in particular it may then be reasonable to assume $w_X = 1$ for all $X$ in $\mathcal{P}$.

Typically, most of the background information $E_o$, for example information about alibis or eye-witness reports, has no effect on the likelihood ratio $R_X(E_d|E_o)$. Note, however, that background information about the ethnic groups of $X$ and $Q$, or the relatedness of $X$ with $Q$, can be very important in calculating LRs for DNA evidence.

The weight-of-evidence formula requires modifications in some settings, such as when the crime sample has more than one source (Section 5.2). Nevertheless the formula is very general, and embodies the "in principle" solution to the problem of interpreting DNA profile evidence, including the role of the non-DNA evidence, the effect of relatives, population variability, and laboratory error. By "in principle" we mean that it points out the quantities that need to be assessed and how they should be combined.

One important feature of (8) in connection with DNA evidence is that it provides a demarcation of the roles of jurors and expert witnesses. Ultimately, it is for jurors to assess evidential weight, but (8) indicates that a DNA expert can be most helpful to clear-thinking jurors by guiding them with reasonable values for the $R_X$. The $w_X$ reflect jurors' assessments of the non-DNA evidence, and will not usually be a matter for the (DNA expert) forensic scientist.

In the next section we consider various consequences of the weight-of-evidence formula for assessing DNA evidence. We continue to assume that the $R_X$ are given; we defer computing LRs until Section 4.

## 2.2 Consequences for DNA evidence

### 2.2.1 Many possible culprits

Because DNA evidence is widely, and correctly, perceived as being very strong, cases often arise in which there is little or no evidence against the defendant other than the DNA evidence. In such cases, there may be large numbers of individuals who, if not for that evidence, would be just as likely to be the culprit as the defendant (in other words, many individuals $X$ for whom $w_X$ is not small).

**The weight-of-evidence fallacy: examples**

Many commentators, seem to take the view that the fact that the profile is rare (i.e. $p$ is small) alone establishes guilt. Some examples of statements that seem to be based on this fallacy are:

- "There is absolutely no need to come in with figures like 'one in a billion', 'one in ten thousand' is just as good".
- "population frequencies ... $10^{-5}$ or $10^{-7}$. The distinction is irrelevant for courtroom use"

These statements are misleading because in the presence of many possible culprits, or strong exculpatory evidence, very small LRs may be consistent with acquittal, and differences of one or two orders of magnitude may be crucial.

Even if all the LRs are very small, this may not suffice to imply a high probability for the defendant's guilt since the bottom line of the weight-of-evidence formula (8) involves a summation, and the total of many small quantities may not be small. A juror told only that 1 in 1 million persons has this profile may incorrectly conclude that this amounts to overwhelming proof of the defendant's guilt. This error can be extremely detrimental to defendants when there are many alternative possible culprits, or substantial exculpatory evidence.

### 2.2.2 Incorporating the non-DNA evidence

The overall cases against the defendant, $Q$, in the two assault cases (see box) differ dramatically: in the first case the evidence against $Q$ seems overwhelming; in the second, a jury would have to make careful judgements about the validity of the alibi, the possibility of travelling such a distance, and the strength of the DNA evidence. In the weight-of-evidence formula, the difference between these two cases is encapsulated in different values

## Two assault cases

Case 1
- victim recognises alleged assailant and reports his name, $Q$, to police;
- $Q$ is found to have injuries consistent with the victim's allegation and cannot give a convincing alibi for his whereabouts at the time of the alleged offence;
- $Q$ is profiled and found to match the crime profile.

Case 2
- victim does not see assailant and can give no useful information about him;
- the crime profile is compared with DNA profiles from many other individuals until a matching individual $Q$ is found.
- $Q$ lives in another part of the country, has a good alibi for the time of the crime and no additional evidence can be found linking him to the alleged offence.

for the $w_X$. A plausible allegation by the victim in Case 1 may lead a juror to assign small values of $w_X$ to each alternative possible culprit $X$. Lacking such an allegation, and faced with strong alibi evidence, jurors may assign values greater than one to many of the $w_X$.

A juror may be reluctant to assign precise values to the $w_X$, but can make broad distinctions between the moderately large and extremely small values that may be appropriate in these two examples.

### 2.2.3 Relatives

Because DNA profiles are inherited, closely related individuals are more likely to share a DNA profile than are unrelated individuals. Many commentators have taken the view that close relatives of the defendant need not be considered unless there is specific evidence to cast suspicion on them.

The weight-of-evidence formula shows this view to be mistaken. Consider the case outlined in the box below. It may be helpful to profile brothers in such cases, if possible. The brother may, however, be missing, or refuse to co-operate. It may not even be known whether or not the defendant has any brothers.

The probability of the defendant's innocence in this case is about 1%. A juror may or may not choose to convict on the basis of this calculation: the pertinent point is that ignoring the brother would give a very misleading view of evidential strength, leading to probability of innocence of only 0·01%. It is easy to think of similar situations, involving additional unexcluded brothers or other close relatives, in which the probability of innocence is substantial, even after apparently strong DNA evidence has been taken into

account.

## Relatives calculation: an example

There is direct DNA profile evidence against a defendant $Q$, but the DNA profiles of the other possible culprits – a brother of $Q$ named $B$ and 100 unrelated men – are not available. The non-DNA evidence does not distinguish between these 102 individuals, so that the "other evidence" ratios $w_X$ are all equal to one. We will defer consideration of methods for calculating LRs until Section 4. Here, we will take the following values: for the brother, $R_B = 1/100$, for all other possible culprits $R_X = 1/1\,000\,000$. Then

$$\mathrm{P}(G|E) = \frac{1}{1 + 1/100 + 100/1\,000\,000} \approx 99\%, \tag{10}$$

Although still more simple than realistic cases, the example serves to illustrate the general point that consideration of unexcluded close relatives may be enough to raise reasonable doubt about the defendant's guilt *even when there is no direct evidence to cast suspicion on the relatives*.

### 2.2.4 Laboratory and handling errors

If crime and defendant profiles originate from the same individual, the observation of matching profiles is not surprising. Non-matching profiles could nevertheless have arisen through an error in the laboratory or at the crime scene, such as an incorrect sample label or laboratory record, a contaminated sample, a software error in a computer-driven laboratory procedure or, possibly, tampering with evidence. The common practice of ignoring this possibility favours the defendant, although the effect is typically small.

On the other hand, when the defendant is not guilty, ignoring the possibility of error is always detrimental to the defendant, sometimes substantially so. The observed match could have arisen in two ways:

(a) suspect and culprit happen to have matching DNA profiles and no typing error occurred;

(b) suspect and culprit have distinct DNA profiles, and the observation of matching profiles is due to an error in one or both recorded profiles.

Both (a) and (b) are typically unlikely. In many cases (b) may be important, but (a) may be the focus of more attention, in part because error probabilities are difficult to assess. Even if error rates from external, blind trials are available, there will usually be specific

details of the case at hand that differ from the circumstances under which the trials were conducted, and that make it more or less likely that an error has occurred.

We saw in Sections 1.3.2 and 2.4.2 the role of error probabilities under very simple assumptions. Some broad conclusions of these analyses are:

- In order to achieve a satisfactory conviction based primarily on DNA evidence, the prosecution needs to persuade the jury that the relevant error probabilities are small.
- If the probability of error (b) is much greater than the probability of matching profiles (a), then the LR corresponding to (a) is effectively irrelevant to evidential weight.
- What matters are not the probabilities of *any* profiling or handling errors, but only the probabilities of errors that could have led to the observed DNA profile match.

## 2.3 Database searches

The UK has a national database of the DNA profiles of named individuals for criminal intelligence purposes. At 31/03/16, over 5 million people in the UK had their profiles recorded in the UK NDNAD (80% men, 20% women; 78% white, 7% black, 5% South Asian, 8% unknown), while the number of profiles from unsolved crimes was 0.5 million. In 2014/15, the NDNAD produced 220 subject to crime scene matches from an urgent search of the NDNAD, including to 43 homicides and 72 rapes. It also produced 29,315 routine subject to crime scene matches, including to 438 homicides and 635 rapes. It also provided 1,015 crime scene to crime scene matches and 2,011 partial matches.[1]

The question thus arises as to the appropriate method for assessing the DNA profile evidence when the defendant was identified following a search through this database. The number of individuals involved in such a search, and even the fact that there was a search, may not be reported to the court. This is because intelligence databases consist primarily of the DNA profiles of previous offenders, and admitting that such a search has been conducted is thus tantamount to admitting previous convictions. It is important to know whether or not omitting this information tends to favour the prosecution.

In Section 1.3.3, we considered the related problem of a sequential search in the population of possible offenders in the setting of the island problem. As we discussed there, it is widely – but wrongly – believed that the fact that a DNA profile match is more likely when it results from a search means that the evidence is weakened by the search.

The correct analysis shows that Lesson 4 (page 12) still holds, and DNA evidence is usually slightly stronger in the database search setting than when no search has occurred. Omitting information about the search tends to favour the defendant, although usually the effect is small. This analysis requires a modification of the weight-of-evidence formula,

[1] NDNAD annual report 2012 to 2013, available at `https://www.gov.uk/government/publications/national-dna-database-annual-report-2014-to-2015`

to take into account the fact that individuals have been observed not to match the crime DNA profile. The intuition behind the database search result is two-fold:

(a) the other individuals in the database were found not to match and hence are effectively excluded from suspicion, reducing the number of possible culprits;

(b) the observation of many non-matches strengthens the belief that the profile is rare.

As an illustration of (a), consider an enormous database that records the DNA profiles of everyone on earth. If the defendant's profile were the only one in this database to match the crime profile, then the evidence against him would clearly be overwhelming.

Although the DNA evidence may be slightly stronger in the context of a database search, the overall case against the defendant may tend to be weaker because there may often be little or no non-DNA evidence against the defendant.

## 2.4 Derivation of the weight-of-evidence formula

So far I have stated many results without derivation. In this section I fill in some of the missing details.

### 2.4.1 Bayes Theorem

Given evidence $E$ and the two hypotheses confronting a criminal juror:

$H_Q$: $Q$ is guilty, and

$I$: $Q$ is not guilty,

Bayes Theorem describes how to update *prior* probabilities of $H_Q$ and $I$ to take into account of the information conveyed by $E$. Since exactly one of $H_Q$ and $I$ is true, we must have $\mathrm{P}(H_Q) + \mathrm{P}(I) = 1$. Bayes Theorem is

$$\mathrm{P}(H_Q|E) = \frac{\mathrm{P}(E|H_Q)\mathrm{P}(H_Q)}{\mathrm{P}(E|H_Q)\mathrm{P}(H_Q) + \mathrm{P}(E|I)\mathrm{P}(I)}. \tag{11}$$

All the probabilities in (11) are conditional on background information $E_o$.

Although valid, (11) is not immediately useful for DNA evidence because the likelihood $\mathrm{P}(E|I)$ cannot be directly calculated. If $Q$ is the source of the crime scene DNA (which for now we assume is equivalent to guilt), then the probability $\mathrm{P}(E|H_Q)$ of observing the DNA evidence is relatively straightforward. If $Q$ isn't the source, however, we cannot evaluate the probability of the DNA evidence without knowing something about the person who was the source.

To overcome this problem it is convenient to partition the event $X$ into a union of events $H_X$, where $X$ denotes an individual other than $Q$. Then $\mathrm{P}(I) = \sum_X \mathrm{P}(H_X)$ and

$\mathrm{P}(E|I)\mathrm{P}(I) = \sum_X \mathrm{P}(E|H_X)\mathrm{P}(H_X)$. Substitution in (11), leads to the weight-of-evidence formula, (8).

Replacing $I$ with $\cup_X\{H_X\}$ is convenient, but is not appropriate in every forensic identification setting. One exception arises when there are multiple contributors to the crime scene DNA profile (Section 5.2) in which case the alternative hypotheses must each specify all of the contributors of DNA.

### 2.4.2 Typing errors

Consider again the modification to the island problem discussed in Section 1.3.2, in which typing errors occur independently with probabilities $\epsilon_1$ and $\epsilon_2$. Here, if suspect and culprit are not the same person, then the evidence must have arisen in one of three ways:

- Both suspect and culprit have $\Upsilon$, and no typing error occurred; this has probability $p^2(1-\epsilon_2)^2$.
- One of the two has $\Upsilon$, the other does not but a false positive error occurred; this has probability $2p(1-p)\epsilon_1(1-\epsilon_2)$.
- Neither suspect nor culprit have $\Upsilon$, and both were incorrectly typed; this has probability $(1-p)^2\epsilon_1^2$.

If suspect and culprit are the same person, then there are two ways to have observed the evidence:

- The suspect/culprit has $\Upsilon$ and was correctly typed twice; this has probability $p(1-\epsilon_2)^2$.
- The suspect/culprit does not have $\Upsilon$ and was incorrectly typed twice; this has probability $(1-p)\epsilon_1^2$.

Combining all these probabilities we obtain

$$\begin{aligned} R_X &= \frac{p^2(1-\epsilon_2)^2 + 2p\epsilon_1(1-p)(1-\epsilon_2) + (1-p)^2\epsilon_1^2}{p(1-\epsilon_2)^2 + (1-p)\epsilon_1^2} \\ &= \frac{(p + \epsilon_1 - p(\epsilon_1+\epsilon_2))^2}{p(1-\epsilon_2)^2 + (1-p)\epsilon_1^2} \\ &\approx \frac{(p+\epsilon_1)^2}{p}, \end{aligned}$$

The final approximation holds if $p$, $\epsilon_1$ and $\epsilon_2$ are all small.

# 3 Other approaches to weight of evidence

We have seen that the likelihood ratio approach is very powerful. No matter what unusual circumstances arise in a new case: identical twins, inbreeding, fraud or missing bands, the LR provides us with a framework for assessing the evidence. For every piece of evidence, a juror should ask two questions

- How likely is the evidence if the defendant $Q$ is guilty?
- How likely is the evidence if another individual $X$ is guilty?

Consequently, expert witnesses should provide as much information as possible to help jurors to answer these questions.

Although elegant and powerful, the weight-of-evidence theory based on LRs is often viewed as complicated and unfamiliar. Real crime cases are complicated, so to some extent it is inevitable that a satisfactory theory of evidential weight cannot be very simple. We briefly introduce alternative approaches that seem simpler but that have difficulties.

## 3.1 Uniqueness

Match probabilities for alternative possible culprits unrelated to the defendant $Q$ are often extremely small: for the 10-locus STR system widely used in the UK, calculated match probabilities are usually substantially less than 1 in 1 billion. When a forensic scientist reports match probabilities this small, it seems effectively equivalent to saying that he or she is reasonably certain that $Q$'s DNA profile is unique in the population of possible sources of the crime stain. If so, wouldn't jurors be better assisted by the expert giving a "plain English" statement of this, rather than a match probability whose unfamiliar magnitude may overwhelm or confuse? For example, perhaps an expert witness could assert that, excluding identical twins and laboratory/handling errors, in his/her opinion $Q$'s DNA profile is almost certainly unique in the UK.

Although attractive in some respects, a practice of declaring uniqueness in court does lead to difficulties. One of these is: how to deal with the minority of cases in which uniqueness cannot reasonably be asserted? These often arise for low-template (LTDNA) and/or mixed profiles. Perhaps the most important barrier to declaring uniqueness is the problem of the non-DNA evidence in a case. The event that a particular DNA profile is unique is either true or false, no "objective" probability can be assigned to it. Nevertheless, since this truth or falsity cannot be established in practice, a probability of uniqueness based on the information available to an expert witness, such as that obtained from DNA profile databases, together with population genetics theory, may potentially be useful to a court. The problem then arises as to what data and theory the expert should take into account. Specifically, the non-DNA evidence in a case may be directly relevant, yet it may not be appropriate for the DNA expert to assess this evidence.

Consider a crime scene DNA profile that is thought to be so rare that an expert might be prepared to assert that it is unique. Suppose that, for reasons unrelated to the crime, it

is subsequently noticed that the crime scene profile (CSP) matches that of the Archbishop of Canterbury. On further investigation, it is found to be a matter of public record that the Archbishop was taking tea with the Queen at the time of the offence in another part of the country. A reasonable expert would, in the light of these facts, revise downwards any previous assessment of the probability that the CSP was unique. However, this is just an extreme case of the more general phenomenon that any evidence in favour of a defendant's claim that he is not the source of the crime stain is evidence against the uniqueness of his DNA profile.

Under certain assumptions, including allowing for an $F_{ST}$ value of 2%, simulations indicate that 10 STR loci usually suffice to achieve a 99·9% probability of uniqueness, and 11 loci suffice almost always (Balding, 1999). However, these calculations are based on the crucial assumption that $w_X \leq 1$, which effectively implies that there is no evidence in favour of $Q$. There is sometimes evidence favouring the defendant, and it is not appropriate for the forensic scientist to pre-empt the jurors' assessment of the non-scientific evidence.

Focussing on the directly relevant issue, whether or not $Q$ is the source of the crime stain, rather than uniqueness, makes more efficient use of the evidence and, properly presented and explained to the court, can suffice as a basis for satisfactory prosecutions. A calculation of the probability of "uniqueness" may also provide useful information for courts, provided that a satisfactory way is found to explain the underlying assumptions.

## 3.2 Random Man Not Excluded (RMNE) probabilities

The concept of a "random man" has at least two meanings:

- informally, it means something like "nobody in particular" or "it could be anyone";
- in scientific usage, it means chosen according to a randomising device, such as a computer random-number generator.

Use of the idea of "random man" in the first sense is generally harmless, though may cause some confusion. Serious errors can arise in the weight-of-evidence setting when the term is used in the second sense. It is important to remember that "random man" doesn't exist in real crime cases: nobody is actually chosen at random, and so probabilities calculated under an assumption of randomly sampled suspects can have no direct bearing on evidential weight.

One example of the errors that "random man" can cause concerns the argument over which population the man is supposed to have been randomly drawn from, that was pervasive in the forensic DNA scientific literature during the early 1990s. These arguments can never be resolved. More importantly, the "random man" concept ignores the differing levels of relatedness between defendant and other possible culprits, and often leads people to ignore the role of the *number* of possible culprits in evidential assessments. Clear thinking about typing errors and the effect of searches can also be undermined.

The concept of "random man" underlies the *inclusion probability*, also called the RMNE probability, for the profiling tests that were performed. This probability may be reported to a court instead of an LR. For example, consider a paternity testing scenario in which a child's paternal allele at a locus has population relative frequency $p$, and we wish to compare the hypothesis "$X$ is the father" with " $Q$ is the father", where the genotype of $Q$ at the locus is available but that of $X$ is not. In simplified settings (see Section 5.1), the LR is $2p$ if $Q$ has just one copy of the child's paternal allele, and $p$ if $Q$ is homozygous for that allele. Under simple assumptions, the probability that a random man would not have a copy of the child's paternal allele is $(1-p)^2$, and so the RMNE probability is $2p - p^2$, irrespective of the genotype of $Q$.

The RMNE probability seems attractive as a measure of evidential weight, but there is no theory linking it with the question of the defendant's guilt. This, in itself, may not be troubling if it satisfied some informal notion of fairness, and this is the case in many settings but, unfortunately, the RMNE probability does lead astray in some cases. Their key weakness is that it does not involve the profile of the defendant. Consequently:

- All evidence decreases the RMNE probability and hence counts *against* the defendant. This means that test results that actually favour the defendant (i.e. that decrease the probability of guilt) will wrongly count against him/her in the RMNE approach. In the paternity testing scenario outlined above, if $Q$ is found to have only one copy of the child's paternal allele then if $p > 0{\cdot}5$ he thereby has a *reduced* probability of being the father (see Section 5.1). The RMNE approach wrongly counts this as evidence against $Q$, just the same as if he had two copies of the child's paternal allele. Although $p > 0{\cdot}5$ is rarely realistic for today's DNA profiling systems, this scenario highlights the logical problems associated with not answering the relevant question.

- In a case involving two co-defendants both of whom are said to have contributed DNA to the sample, the evidence can weigh more heavily against one defendant than the other, whereas the RMNE probability will be the same for both defendants.

In single-contributor identification cases "inclusion" corresponds to having the CSP, and hence the RMNE approach is similar to the LR approach except that the effect of shared ancestry between $Q$ and $X$ cannot readily be taken into account. For LTDNA profiles, the RMNE probability faces severe difficulties. In general, it may have some uses in measuring and conveying evidential weight, but it should not be used without checking that it does not conflict with the logical analysis based on LRs.

## 3.3 Hypothesis Testing

In science, weight-of-evidence is often assessed via significance levels and/or $p$-values. The jury in a criminal case must reason from the evidence presented to it, to a decision between

the hypotheses:

$$\begin{aligned} H_Q &: Q \text{ is guilty;} \\ I &: Q \text{ is not guilty.} \end{aligned}$$

Within the hypothesis-testing framework, the legal maxim "innocent until proven guilty" would imply that $I$ should be the null hypothesis, and so the probability of a match under $I$ can be interpreted as a $p$-value.

But how do we calculate a $p$-value taking into account the possibility that the true culprit could be a relative of the defendant? The usual answer is to invoke "random man", and assume that hypothesis $I$ implies that $Q$ has been chosen randomly in some population of innocent suspects. But since no random sampling really took place, it is impossible to specify the population. Too broad a definition of the population leads to overstatement of the evidence, because a large population must contains many people sharing little ancestry with $Q$. If we try to avoid this overstatement by specifying the narrowest possible population, we are led to the population consisting of $Q$ only, in which the match probability is one.

The hypothesis testing framework faces further difficulties with complications that we have seen are readily handled using the weight-of-evidence formula (8):

- How can the $p$-value, assessing the DNA evidence, be incorporated with the other evidence? What if the defendant produces an apparently watertight alibi? What if more incriminating evidence is found?
- How should the possibility of laboratory or handling error be taken into account?
- What if the defendant was identified only after a number of other possible culprits were investigated and found not to match?

Perhaps the most important weakness is the first: the problem of incorporating the DNA evidence with the other evidence. Hypothesis tests are designed to make accept/reject decisions on the basis of the scientific evidence only, irrespective of the other evidence. Legal decision makers must synthesise all the evidence, much of it unscientific and difficult to quantify, in order to arrive at a verdict.

# 4 Calculating LRs allowing for coancestry

## 4.1 Some population genetics

### 4.1.1 Population genotype probabilities

An individual's genotype at an STR locus usually consists of two alleles, one paternal and one maternal in origin. Alleles are conventionally labelled according to the number of repeat units, so that a genotype might be represented by the unordered allele pair 7,9

(unordered because we cannot usually say which allele is paternal and which is maternal). If the individual has the same allele, say the 11, from both parents, the genotype would be represented as 11,11. Below we will use $A$, $B$, and $C$ for arbitrary STR alleles, and we will write their population allele fractions as $p_A$, $p_B$, and $p_C$.

These fractions are unknown, and indeed unknowable because the relevant population is not well defined. However, population allele fractions are routinely estimated in several loosely-defined population groups, usually based on ethnic appearance as assessed by a police officer. In the UK, the most important ethnic groups for forensic DNA profiles are: Caucasians (which can include people of West Asian or North African origin), Afro-Caribbeans (which can include sub-Saharan Africans), and South Asians (also called Indo-Pakistani). Clearly, the classification of the UK population into these groups is arbitrary, many individuals do not fit well into any of them, and in particular the Afro-Caribbean group is genetically heterogeneous.

Allele fractions in these groups are estimated from samples whose size is typically a few hundreds (Figure 1). Allowance for the effects of sampling error is briefly discussed below in Section 4.3.1. However, in addition to the problem of defining the ethnic groups, the samples used to estimate allele fractions in them are not scientific, random samples, but are "convenience" samples whose representativeness is unknown. Figure 1 shows that the allele frequencies do not differ dramatically over these groups, which is encouraging that the effects of ethnic misclassification and non-representative samples will typically not be great, and that use of a sufficiently large $F_{ST}$ value (Section 4.3.2) can compensate for these effects.

A common assumption in human population genetics is that an individual's maternal and paternal alleles can be regarded as independent. This is known as "Hardy-Weinberg equilibrium", but this is a misleading term as the relevant issue is independence and not equilibrium. Possible reasons for non-independence include inbreeding, selection, and genotyping error, but independence appears to hold to a good approximation at most genomic loci in most human populations. Under independence, and given the population allele fractions, the genotype fractions in the heterozygote ($AB$) and homozygote ($AA$) case, conditional on the allele fractions, are:

| genotype: | $AB$ | $AA$ |
|---|---|---|
| population fraction: | $2p_Ap_B$ | $p_A^2$ |

### 4.1.2 A sampling formula for alleles

Each of the major ethnic groups can be regarded as being composed of many subpopulations, based for example on geographical boundaries, recent migration, or religious or other social groupings. If an allele has average fraction $p$ over the subpopulations, its variance can be written in the form $F_{ST}p(1-p)$, where $F_{ST}$ is a population genetics parameter between 0 and 1, called the *coancestry coefficient*. $F_{ST}$ measures the relatedness among individuals within sub-populations relative to the total population (Figure 2). More relatedness within subpopulations, relative to that in the total population, means higher $F_{ST}$

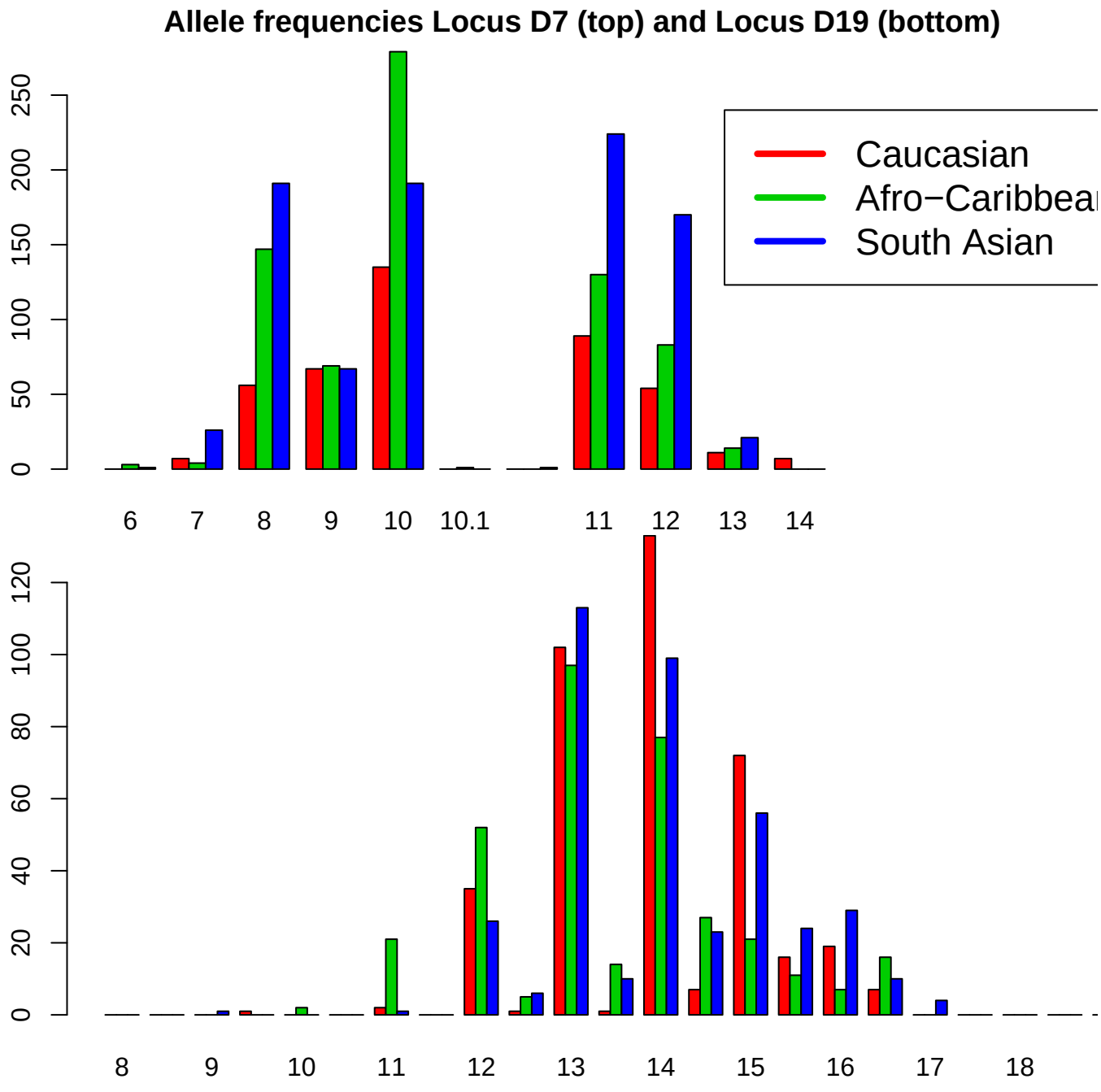



Figure 1: Allele frequencies in samples from three UK population groups at two STR loci.

and a greater variation in allele fractions across subpopulations. This has implications for the LR associated with a DNA profile match, since if $Q$ has some relatedness with an alternative possible source of the DNA, a match becomes more likely.

There is a sampling formula for DNA alleles that can take this effect into account. Suppose that $n$ alleles have been sampled in the subpopulation, of which $n_A$ are $A$. Then the probability that the next allele sampled is also $A$ can be written as:

$$\frac{n_A F_{ST} + (1-F_{ST})p_A}{1 + (n-1)F_{ST}}. \tag{12}$$

When $n_A = n = 0$, we obtain probability $p_A$ that the first allele drawn is $A$. The probability that the first two alleles drawn are both $A$ is

$$p_A(F_{ST} + (1-F_{ST})p_A) = p_A^2 + F_{ST}p_A(1-p_A). \tag{13}$$

Roughly speaking, increasing $F_{ST}$ increases the probability of observing two $A$ alleles, because the first observation of an $A$ allele suggests that they are relatively common in the subpopulation, and so drawing another $A$ is less surprising. If there were no subpopulation variation, the second $A$ allele would have the same probability as the first, and so the probability of two $A$ alleles is $p_A^2$, obtained by substituting $F_{ST} = 0$ in (13). Increasing $F_{ST}$ also increases the probability of two $B$ alleles, but decreases the probability of an $A$ allele followed by a $B$, which is:

$$(1-F_{ST})p_A p_B. \tag{14}$$

Formula (12) can be used to build up probabilities for larger samples of alleles from a subpopulation. Samples of size four have a special importance in forensic identification problems, because of the two alleles at each locus from each of $Q$ and $X$. Using (12), the probability of observing $ABAB$ (in that order) in a sample of size four is:

$$\frac{p_A p_B(1-F_{ST})(F_{ST}+(1-F_{ST})p_A)(F_{ST}+(1-F_{ST})p_B)}{(1+F_{ST})(1+2F_{ST})}, \tag{15}$$

while the probability of $AAAA$ is

$$\frac{p_A(F_{ST}+(1-F_{ST})p_A)(2F_{ST}+(1-F_{ST})p_A)(3F_{ST}+(1-F_{ST})p_A)}{(1+F_{ST})(1+2F_{ST})}. \tag{16}$$

As $F_{ST}$ increases, the probability of $AAAA$ rises up to $p_A$.

## 4.2 Identification: single contributor

We need to compute $R_X$, the ratio of the probability of the evidence $E$ if $X$ is the source of the DNA, to its probability if the defendant $Q$ was the source. We will assume that $E$ consists of the information that both CSP $= G$ and $\mathcal{G}_Q = G$, for some genotype $G$, where

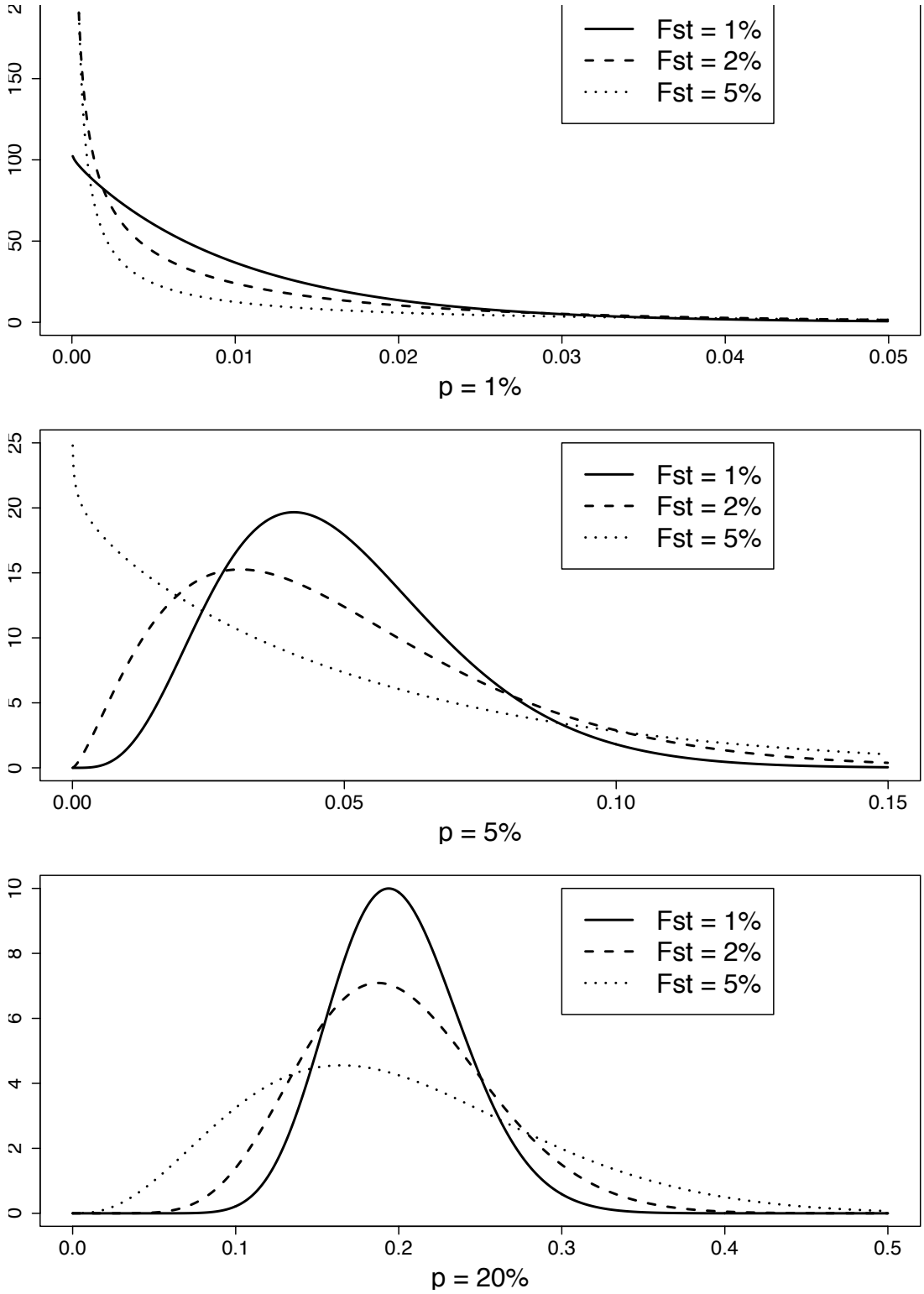



Figure 2: Distributions of subpopulation allele fractions for given values of $F_{ST}$ and the global allele fraction $p$.

CSP = Crime Scene Profile and we introduce the notation $\mathcal{G}_y$ for the genotype of person $y$. Then

$$R_X = \frac{\mathrm{P}(\mathrm{CSP}{=}\mathcal{G}_Q{=}G|H_X)}{\mathrm{P}(\mathrm{CSP}{=}\mathcal{G}_Q{=}G|H_Q)}. \tag{17}$$

Ignoring the possibility of error, CSP and $\mathcal{G}_Q$ are equivalent under $H_Q$, whereas under $H_X$ the CSP informs us about $\mathcal{G}_Q$. So (17) can be simplified further to

$$R_X = \frac{\mathrm{P}(\mathcal{G}_X{=}\mathcal{G}_Q{=}G)}{\mathrm{P}(\mathcal{G}_Q{=}G)} = \mathrm{P}(\mathcal{G}_X{=}G|\mathcal{G}_Q{=}G). \tag{18}$$

Thus, $R_X$ is the conditional probability, called the "match probability", that $X$ has genotype $G$ given that $Q$ has it. Population genetic effects arise, and can be dealt with, via this conditioning. The important feature of the match probability is that it takes account of *both* the observed profiles that form the match. Some authors misleadingly refer to the population relative frequency of the profile as a "match probability" which is inappropriate since the concept of "match" involves two profiles, rather than just one.

### 4.2.1 Single-locus match probabilities

Suppose that both $X$ and $Q$ are homozygous for allele $A$. If we assume that they are unrelated, both come from the same subpopulation, and neither is inbred, then the numerator of (18) corresponds to the probability (16) that a sample of four alleles from the subpopulation is $AAAA$. Dividing by the probability (13) that a sample of two alleles is $AA$, we obtain:

**Single locus match probability: CSP $= \mathcal{G}_Q = AA$**

$$\frac{(2F_{ST}{+}(1{-}F_{ST})p_A)(3F_{ST}{+}(1{-}F_{ST})p_A)}{(1+F_{ST})(1+2F_{ST})} \tag{19}$$

In the heterozygous case, under the same assumptions, we need the probability that, in a sample of size four, the first and second pairs of alleles are both $AB$, divided by the probability that a sample of size two is $AB$. These probabilities are given at ((15) and (14), except that we have to multiply by 2 for each pair of alleles (because of the two possible allele orderings), which gives:

These match probabilities are shown in Figure 3 for $F_{ST}$ ranging from 0 to 1. Notice that increasing $F_{ST}$ does not necessarily increase the match probability in the heterozygous case, although this almost always occurs in practice. Whatever the values of $p_A$ and $p_B$, as $F_{ST}$ approaches 1, the homozygous and heterozygous match probabilities approach 1 and 1/3, respectively.

**Single locus match probability: CSP $= \mathcal{G}_Q = AB$**

$$2\frac{(F_{ST}+(1-F_{ST})p_A)(F_{ST}+(1-F_{ST})p_B)}{(1+F_{ST})(1+2F_{ST})} \tag{20}$$

Although $F_{ST}$ cannot encapsulate all population genetics phenomena, it does capture the essentials relevant to forensic match probabilities. Selection and other phenomena can distort population profile frequencies away from estimates based on assuming independence across loci, but there is no reason to expect these to systematically favour or disfavour defendants. The only population genetics phenomenon that, if ignored, systematically disfavours defendants is shared ancestry between defendant and alternative possible culprits – which is what $F_{ST}$ accounts for. Human population genetics is complicated, and inevitably $F_{ST}$ is an imperfect measure, but by choosing a sufficiently large value defendants will not be systematically disfavoured, while match probabilities remain small enough to form the basis of satisfactory prosecutions in most cases.

### 4.2.2 Multiple loci: the "product rule"

Match probabilities at multiple loci can be obtained by multiplying together the match probabilities at the individual loci calculated using (19) and (20). This implies an assumption of independence across loci, which is reasonable when the main source of dependence, which is coancestry, has been accounted for by a sufficiently large value of $F_{ST}$. The product rule was for a long time controversial, and like many controversies the debate was marred by confusion over definitions. Whether or not the alleles at different loci are independent is not an absolute state of nature, but depends on what information has been taken into account. If relatedness and coancestry have not been taken into account, then an assumption of independence is clearly unfair to defendants, but this problem can be eliminated by adjusting for $F_{ST}$. In both cases match probabilities are multiplied over loci and so both may be described as applications of the "product rule".

### 4.2.3 Relatives of $Q$

Let $Z$ denote the number of alleles at a locus that $X$ and $Q$ share directly from a known, recent common ancestor (e.g. parent or grandparent). We assume here that only the DNA profile of $Q$ is available to the court. The probability distribution for $Z$ under some common, regular relationships (i.e. no inbreeding) are shown in Table 1.

If $Z = 0$, we are in the situation of Section 4.2.1 above ($Q$ and $X$ unrelated) and we write $M_2$ for the appropriate match probability, either (19) or (20). If $Z = 2$, a match is certain. For $Z = 1$, consider first the case of an $AA$ homozygote match. Although there is a total of four alleles at the locus, two of these are shared from the recent common ancestor

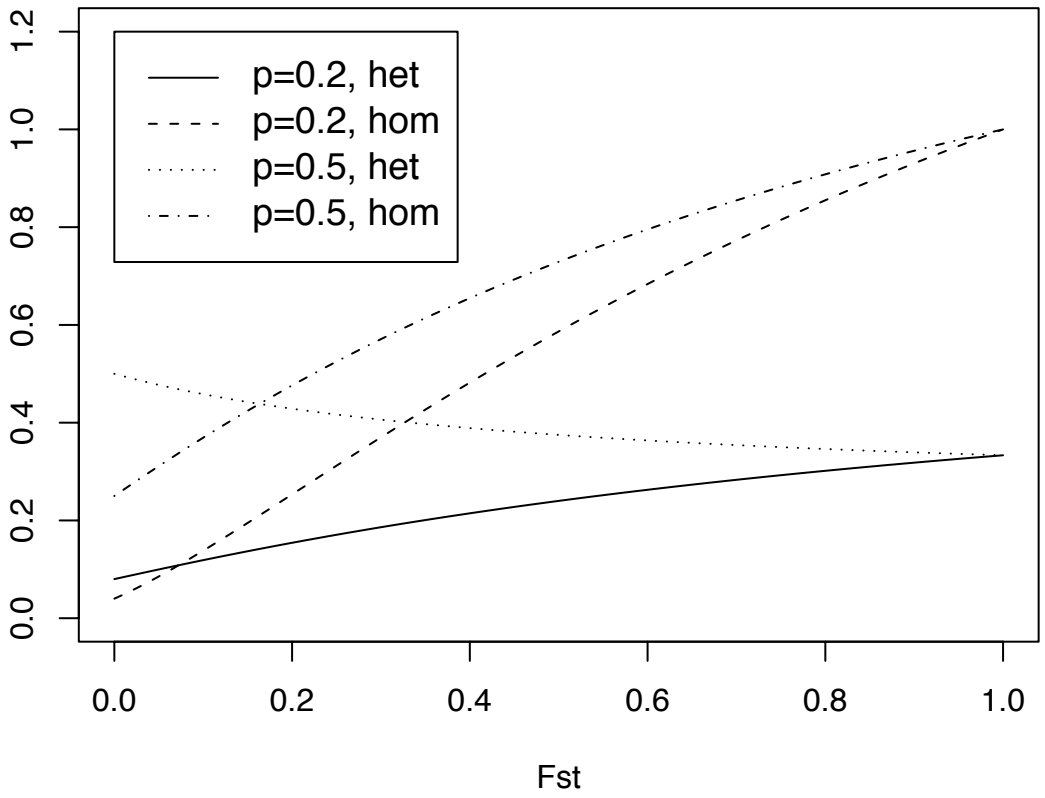



Figure 3: Heterozygote and homozygote single-locus match probabilities as a function of $F_{ST}$ for two values of $p$. In the heterozygote case, the fractions of the two alleles are both equal to $p$.

and so effectively we have observed three matching alleles. The match probability is the probability of this observation given that the first two alleles match, which using (12) is

$$M_1 = \frac{2F_{ST} + (1-F_{ST})p_A}{1 + F_{ST}}.$$

In the heterozygous case, there are two equally-likely possibilities ($A$ or $B$) for the allele that is shared from the recent common ancestor, and the match probability is the probability that the third allele sampled is the other one of $A$ and $B$. Thus

$$M_1 = \frac{F_{ST} + (1-F_{ST})(p_A+p_B)/2}{1 + F_{ST}}.$$

The overall match probability for relatives is then

$$\mathrm{P}(Z{=}2) + M_1\mathrm{P}(Z{=}1) + M_2\mathrm{P}(Z{=}0).$$

| Relationship | P($Z$=0) | P($Z$=1) | P($Z$=2) |
|---|---|---|---|
| Sibling | 1/4 | 1/2 | 1/4 |
| Parent | 0 | 1 | 0 |
| Half-sib | 1/2 | 1/2 | 0 |
| Cousin | 3/4 | 1/4 | 0 |
| Unrelated | 1 | 0 | 0 |

Table 1: Relatedness coefficients for some regular relatives. The coefficients for aunt/uncle – niece/nephew and grandparent – grandchild are the same as those for a half-sib.

## 4.3 Issues for casework

According to the weight-of-evidence formula (8), a rational juror should to assess the LR (or match probability) for every alternative possible culprit $X$. In the preceding subsections we have derived formulas for match probabilities, in terms of population allele fractions (the $p_j$) and $F_{ST}$. It remains then to choose appropriate values for these parameters for the alternative suspects in a particular case.

Recall that the LR for DNA evidence is the probability of the evidence under $H_X$ divided by its probability under $H_Q$. It follows that in principle the database to be used to estimate the $p_j$ should be that most appropriate for $X$, the alternative possible culprit under consideration. The relevant value of $F_{ST}$ is that which describes the coancestry of $X$ with the defendant $Q$, relative to the population in which the $p_j$ have been estimated. If $X$ is not from the same ethnic group as $Q$, then they have little coancestry and so a small value of $F_{ST}$ can be justified. Larger values are appropriate if $X$ and $Q$ share the same ethnic background.

In practice it may be satisfactory to use some "conservative" values for the $p_j$ and $F_{ST}$ (tending not to overstate the evidence against $Q$), giving a single LR for all $X$ except the direct relatives of $Q$.

### 4.3.1 Values for the $p_j$

In applying (19) and (20) to actual cases, it seems natural to estimate $p_j$, the population proportion of allele $j$, by its relative frequency in the database most relevant to the alternative possible culprit $X$. The database closest to $Q$ will in general provide a conservative alternative for all $X$, and the difference should be small given an appropriate value for $F_{ST}$ (see below).

The sampling uncertainty in the estimate of $p_j$ still needs to be taken into account (Section 1.3.1). One approach to this is to estimate $p$ at a heterozygous locus by $(n_j + 2)/(n+4)$, where $n_j$ is the database count of allele $j$, and $n$ is the number of alleles in the database. This can be thought of as adding both the crime scene and defendant profiles to the database. In the homozygous case, the analogous estimate is $(n_j + 4)/(n + 4)$.

### 4.3.2 The value of $F_{ST}$

Published estimates of $F_{ST}$ at STR loci, for subpopulations of the Caucasian population are often small, and typically less than 1%. There are nevertheless several arguments for using larger values in forensic practice:

- The appropriate value of $F_{ST}$ is never available, and we are reduced to using an educated guess. Among the sources of uncertainty is that subpopulations are never well defined and "Caucasian" is also not well defined. In order to avoid unfairness to the defendant, we should prefer to over-estimate rather than under-estimate $F_{ST}$.

- Strictly, we should integrate the match probability over a distribution of values for $F_{ST}$, and because the integration is a product over many terms the integral will be dominated by the upper tail of the distribution.

- Published estimates of $F_{ST}$ usually relate to the variation of allele fractions around the observed mean value. However in forensic applications, what is needed is the variation of subpopulation values away from the forensic database value, and this may well be substantially larger. In particular, minority ethnic groups may be heterogeneous and the database allele frequency may not be representative of the specific ethnic group relevant to a particular crime.

For these reasons I suggest that a relatively large value, such as 2%, be used in UK forensic practice when both suspect and alternative possible culprit are Caucasians, and perhaps 3% could be used if both are drawn from one of the large minority groups (Afro-Caribbeans or South Asians). In some small minority groups, $F_{ST} = 5\%$ may be appropriate.

### 4.3.3 The hypotheses

The hypotheses of direct interest to the court are

$$Q \text{ is guilty} \qquad \text{and} \qquad Q \text{ is not guilty.}$$

However, it is usually not appropriate for a forensic scientist expert witness to comment directly on these hypotheses. Instead, he or she might reasonably compute LRs comparing hypotheses of the form

$$\begin{aligned} H_Q &: \quad Q \text{ is the source of the crime-scene DNA sample;} \\ H_X &: \quad X \text{ is the source of the crime-scene DNA sample.} \end{aligned}$$

More complex hypotheses may be appropriate in some settings, such as those involving multiple contributors to the crime sample, discussed below in Section 5.2. In some cases the nature of the tissue from which the DNA is obtained (e.g. blood or semen) is relevant to the hypotheses.

In principal, the question of which hypotheses should be compared is one for the courts and not the forensic scientist. However, the work of the forensic scientist is usually

performed in advance of the court sitting. Thus s/he needs to consider all reasonable possibilities for the hypotheses of interest to the court. In particular, choice of defence hypothesis can be problematic, since the defence is not required to propose any specific alternative to the prosecution allegation. One criterion for choosing among a range of possible defence hypotheses is to select that which is most favourable to the defence (gives an LR closest to one). However, it is not reasonable to adhere strictly to this criterion, since for DNA evidence a hypothesis that the source of the DNA was an identical twin of $Q$ would always give LR = 1.

# 5 More complex scenarios

## 5.1 Paternity and relatedness

The general issues for paternity and other relatedness testing are similar to those for identification. In the case that a man $Q$ is alleged to be the father of a child, the relevant quantities to assess are the LRs:

$$R_X = \frac{\mathrm{P}(E|X \text{ is the father})}{\mathrm{P}(E|\ Q \text{ is the father})}, \tag{21}$$

for each alternative possible father, $X$. In (21), $E$ denotes all the evidence, but we will here ignore the non-DNA evidence (the principles are the same as in Section 2.2.2) and assume that $E$ consists only of $(\mathcal{G}_Q, \mathcal{G}_c, \mathcal{G}_m)$: the genotypes of $Q$, the child $c$ and its mother $m$. We will also assume that $X$ is not directly related to either $m$ or $Q$, and we will ignore the possibility of genotyping error or mutation.

### 5.1.1 Single locus: paternal allele known

Assume first that $c$ has exactly one allele in common with $m$, and so $c$'s other allele is paternal in origin and $Q$ is excluded as a possible father if this allele is not included in his genotype. Otherwise, (21) can be expressed as:

$$\frac{\mathrm{P}(c\text{'s paternal allele}|\mathcal{G}_Q, \mathcal{G}_m, X \text{ is father})}{\mathrm{P}(c\text{'s paternal allele}|\mathcal{G}_Q, \mathcal{G}_m, Q \text{ is father})}. \tag{22}$$

Notice that $\mathcal{G}_Q$ and $\mathcal{G}_m$ are now conditioned on (they are to the right of the | in (22)), whereas in the initial formulation (21) they are part of $E$ and hence are to the left of the |. This is similar to $\mathcal{G}_Q$ in the identification setting, and requires a mathematical argument that is omitted here, but broadly speaking any data that are equally likely under both competing hypotheses can be treated in this way, which simplifies calcuations.

The denominator of (22) is 1 if $Q$ is homozygous, and 1/2 otherwise. The numerator is the probability that an allele drawn from $X$ matches $c$'s paternal allele. Ignoring coancestry ($F_{ST} = 0$), this would equal the population fraction of the allele irrespective of $\mathcal{G}_Q$ and $\mathcal{G}_m$, but potential coancestry of $X$ with $Q$ and/or $m$ can alter this value.

If $X$, $Q$, and $m$ are all assumed to have a common level of shared ancestry measured by $F_{ST}$, then we can use the sampling formula (12) to compute (22). Under the hypothesis that $X$ is the father, we have observed five alleles, those of $Q$ and $m$, plus $c$'s paternal allele. If $Q$ is the true father, we have observed four alleles at the locus: those of $Q$ and $m$ ($c$'s alleles aren't counted because they are replicates of observed parental alleles). Thus (22) involves a ratio of two instances of (12), one for a sample of size five and one for a sample of size four.

Alternatively, if $X$ and $Q$ might share ancestry, but $m$ is completely unrelated (for example, from a different ethnic group), then $\mathcal{G}_m$ is uninformative and we should only use $\mathcal{G}_Q$. Thus (22) involves a ratio of two instances of (12), one for a sample of size three and one for a sample of size two.

*Example*

If $\mathcal{G}_c = AC$, $\mathcal{G}_m = AB$, and $\mathcal{G}_Q = AC$, then $c$'s maternal and paternal alleles are, respectively, $A$ and $C$, and (22) becomes:

$$2 \times \mathrm{P}(c \text{ has paternal allele } C | \mathcal{G}_m = AB, \mathcal{G}_Q = AC, X \text{ is father}).$$

Since $X$, $m$ and $Q$ are not directly related, this is the probability that a fifth allele is $C$, given that four alleles have been sampled and found to be $AABC$. If all three individuals are assumed to have a common level of shared ancestry $F_{ST}$, then the sampling formula (12) gives

$$R_X = 2\frac{\mathrm{P}(AABCC)}{\mathrm{P}(AABC)} = 2\left(\frac{F_{ST} + (1-F_{ST})p_C}{1+3F_{ST}}\right). \tag{23}$$

If instead only the shared ancestry of $Q$ and $X$ is taken into account, then

$$R_X = 2\frac{\mathrm{P}(ACC)}{\mathrm{P}(AC)} = 2\left(\frac{F_{ST} + (1-F_{ST})p_C}{1+F_{ST}}\right). \tag{24}$$

If $\mathcal{G}_Q = CD$ instead of $AC$, the LRs (23) and (24) are unchanged. If $\mathcal{G}_Q = CC$, then the factor of two vanishes from both expressions.

Both (23) and (24) reduce to $2p_C$ if shared ancestry is ignored. The effect of shared ancestry is usually to lessen the evidential strength against $Q$, the exceptions arising when $p_C$ is very large (i.e. $C$ is very common).

### 5.1.2 Single locus: paternal allele unknown

Consider for example the LR

$$\frac{\mathrm{P}(\mathcal{G}_c = AB | \mathcal{G}_m = AB, \mathcal{G}_Q = AC, X \text{ is father})}{\mathrm{P}(\mathcal{G}_c = AB | \mathcal{G}_m = AB, \mathcal{G}_Q = AC, Q \text{ is father})}. \tag{25}$$

The denominator is $1/4$, since under the hypothesis that $Q$ is the father the paternal and maternal alleles can be identified, and each has probability $1/2$ given the parental

genotypes. If, however, $X$ is the father, two equally likely possibilities must be considered for the maternal allele, and hence $R_X$ becomes

$$\begin{aligned} 2 \times (&\mathrm{P}(c\text{'s paternal allele is } A | \mathcal{G}_m = AB, \mathcal{G}_Q = AC, X \text{ is father}) + \\ &\mathrm{P}(c\text{'s paternal allele is } B | \mathcal{G}_m = AB, \mathcal{G}_Q = AC, X \text{ is father})). \end{aligned}$$

Assuming a common level of shared ancestry for $X$ and $Q$ only gives:

$$\begin{aligned} R_X &= 2\frac{\mathrm{P}(AAC)}{\mathrm{P}(AC)} + 2\frac{\mathrm{P}(ABC)}{\mathrm{P}(AC)} = 2\left(\frac{F_{ST} + (1-F_{ST})p_A}{1+F_{ST}} + \frac{(1-F_{ST})p_B}{1+F_{ST}}\right) \\ &= 2\left(\frac{F_{ST} + (1-F_{ST})(p_A + p_B)}{1+F_{ST}}\right). \end{aligned}$$

## 5.2 Identification: mixed profiles

### 5.2.1 Visual interpretation of mixed profiles

A mixed profile arises when two or more individuals contribute DNA to a sample. Torres et al. (2003) give a survey of mixtures that have arisen in their own casework. An example of an epg corresponding to a mixed STR profile is shown in Figure 4. The profile at the Amelogenin sex-distinguishing locus (leftmost on the second, "Green", panel) shows a predominance of X, but some trace of Y, suggesting that the mixture stain may have come predominantly from a female, with a minor contribution from a male. This mixture was created in the laboratory, with known contributors, and the above interpretation is indeed correct, but let's proceed for the moment as if we did not know this.

The recorded signals at other loci are consistent with there being two contributors, one "major" and one "minor". For example, at locus VWA (second locus from the left in the top "Blue" panel) the epg indicates the presence of four alleles, two of which (labelled 15 and 18) produce a strong signal, while the remaining two signals (corresponding to alleles 14 and 19) are much weaker, though still clearly distinguishable from the background noise.

Two other loci show similar four-allele patterns in the epg. However, at most of the loci only two or three alleles appear in the epg, which can arise if one or both contributors is homozygous, or if they have alleles in common. At locus D2S1338 (rightmost in the "Blue" panel), there is a strong signal at allele 17, a slightly weaker signal at allele 19, and a much weaker signal at allele 18, suggesting that the major contributor has genotype 17,19 at this locus, while the minor contributor is 17,18. An alternative possibility is that the peak corresponding to allele 18 corresponds to a stutter peak, and only alleles 17 and 19 are actually present in the sample but this is unlikely because the peak, although low, is higher than normal for a stutter peak at this position on the epg. Although the 17,19 and 17,18 seem the most plausible genotype designations for major and minor respectively, it is difficult to assign any measure of confidence to this call.

Interpretation at the TH01 locus (third from left in the "Black" panel) is even more difficult: the two observed alleles, 6 and 9.3, display allele signals of noticeably different

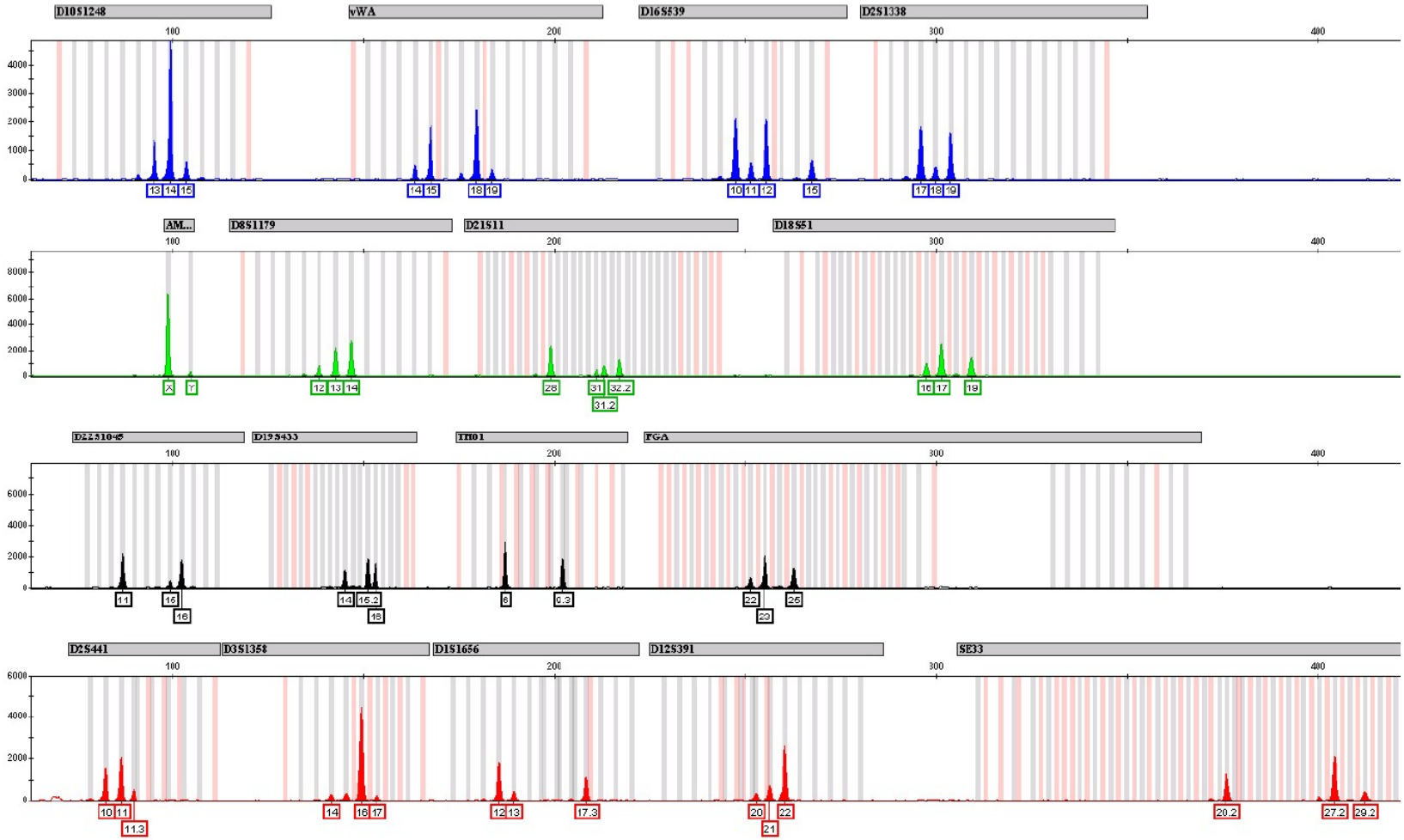



Figure 4: An epg showing a mixed STR profile obtained from a DNA sample with two contributors. The major profile is from a female and the minor component from a male. The dye colours represented in each panel are (from top to bottom): Blue, Green, Black and Red. The DNA sample was amplified with the NGM SElect®STR kit. The amplified fragments were separated on the ABI 3130xL Genetic Analyzer and analysed using the GeneMapper®3.2 software. Image supplied courtesy of Cellmark Forensics. ©2014 Cellmark Forensics.

heights, yet they are not extremely asymmetric. Perhaps the major contributor is 6,9.2 and the minor is a 6,6 homozygote, but other possibilities seem to exist, such as that both contributors are 6,9.3 and the apparent peak imbalance results from a fluctuation in the experimental conditions: here the ratio of peak heights is $\sim 2/3$, which may be regarded as within the normal range of variation of peak heights for a single, heterozygous contributor.

As the above discussion suggests, inferring the profiles of major and minor contributors to a sample can sometimes be done with reasonable confidence, but often it is problematic at least for some loci. The presence of good-quality DNA, and a strong imbalance in the proportions of the DNA from each source individual, facilitate the task. However, in the presence of degraded samples, low DNA copy-number, an unknown number of contributors, or an equal contribution from two contributors, the task can be challenging, and assigning a measure of confidence to any particular genotype designation is problematic.

### 5.2.2 Likelihood ratios under qualitative interpretation

One approach to overcoming the problems with visual interpretation of mixtures, at the cost of discarding the quantitative information from the epg, for example about peak heights and shapes, is to limit interpretation to qualitative allele calling only, without any attempt to infer the underlying genotypes. Thus, the interpretation of locus VWA in the epg of Figure 4 would be limited to the conclusion that alleles 14, 15, and 16 are observed in the mixture. Then, all combinations of underlying genotypes that include at least one copy of each of these alleles are regarded as equally plausible.

We consider here the single-locus case; LRs can be combined across loci via multiplication. These approaches were initially developed by Evett et al. (1991) for non-STR profiles, but they remain applicable to STR profiles. Mortera et al. (2003) describe a probabilistic expert system for the qualitative analysis of DNA mixtures, and the qualitative approach has been extended to take account of the coancestry of all the contributors to a mixture (Curran et al., 1999; Fung and Hu, 2000, 2002) and alternative suspects related to the accused (Fung and Hu, 2004). These authors also offer software for mixture interpretation (Hu and Fung, 2003).

### 5.2.3 The number of unknown contributors

For STR profiles, a minimum number of contributors to a sample of DNA is provided by half the number of distinct alleles observed at any one locus. Even in the case that no more than two alleles are observed at any locus, it is possible that there is more than one contributor although this is typically very unlikely. It is not possible to put an upper bound on the number of contributors, although it may be possible to estimate the number of contributors (Haned et al., 2011; Perez et al., 2011).

In principle, prior to the DNA evidence, judgements about the number of possible sources should lie in the domain of the court, not the DNA expert. However, there will inevitably be occasions when experts make such prior judgements, when they seem

uncontroversial, and when the alternative may be to overly complicate their evidence (e.g. by working out LRs under many different scenarios, most of which are implausible).

We distinguish three classes of contributors to a mixture:

- the contributor of interest, or "contested contributor", who is either $Q$ or $X$ (we will write $Q/X$),
- known contributors, who will be denoted K, K1, K2, ...,
- and unknown contributors denoted U, U1, U2, ....

Although "known" and "unknown" provide a convenient shorthand, the key point is not whether the person is known but whether their reference DNA profile is available for the evaluation. In some mixture CSPs there may be multiple contested contributors, but we propose that those cases be tackled by computing a sequence of LRs each involving only one contested contributor.

### 5.2.4 Two contributors: Q/X and K

The easiest case arises when all the contributors to a mixture other than the contested contributor are known (DNA profiles available). Here we consider the common scenario of one known contributor ($K$) in addition to $Q/X$. Then the evidence is the CSP and the reference profiles of $Q$ and $K$, denoted $\mathcal{G}_Q$ and $\mathcal{G}_K$. It is reasonable to assume that $\mathcal{G}_Q$ and $\mathcal{G}_K$ are equally likely under both hypotheses, and so the LR takes the form

$$R_X = \frac{\text{P(CSP } | \mathcal{G}_Q, \mathcal{G}_K,\ X \text{ and } K \text{ are the sources of the DNA)}}{\text{P(CSP } | \mathcal{G}_Q, \mathcal{G}_K,\ Q \text{ and } K \text{ are the sources of the DNA)}}. \tag{26}$$

At some loci, one or both alleles of $X$ may be masked by the alleles of $K$. In general, the numerator of $R_X$ is computed by summing over the possible genotypes of $X$ given the CSP and $\mathcal{G}_K$. Each term in the sum involves the conditional probability of a particular value of $\mathcal{G}_X$, given $\mathcal{G}_Q$ and $\mathcal{G}_K$ (the conditioning is irrelevant if $F_{ST} = 0$).

*Example*

Consider a single locus, and suppose that the following are observed:

$$\mathcal{G}_Q = \text{AB} \qquad \mathcal{G}_K = \text{AC} \qquad \text{CSP} = \text{ABC}.$$

Then (26) becomes:

$$R_X = \frac{\text{P(ABC} \mid \mathcal{G}_Q = \text{AB}, \mathcal{G}_K = \text{AC}; \text{sources: } X, K)}{\text{P(ABC} \mid \mathcal{G}_Q = \text{AB}, \mathcal{G}_K = \text{AC}; \text{sources: } Q, K)}. \tag{27}$$

The denominator equals one. For the numerator, given that $\mathcal{G}_K = \text{AC}$, the CSP implies that $\mathcal{G}_X$:

- includes a B allele, and
- does not include any allele other than A, B, and C.

The possible genotypes consistent with these two requirements are AB, BB, and BC. Thus, the numerator of (27) is the probability that $X$ has one of these three genotypes, given that $\mathcal{G}_Q = \text{AB}$ and $\mathcal{G}_K = \text{AC}$. Here, we assume:

- that $Q$, $K$, and $X$ are mutually unrelated;
- no coancestry (i.e. $F_{ST} = 0$);
- that the allele proportions $p_A$, $p_B$, and $p_C$ are known; and
- genotypes are in Hardy-Weinberg proportions.

Then (27) becomes:

$$\begin{aligned} R_X &= \text{P}(\mathcal{G}_X = \text{AB}) + \text{P}(\mathcal{G}_X = \text{BB}) + \text{P}(\mathcal{G}_X = \text{BC}) \\ &= 2p_Ap_B + p_B^2 + 2p_Bp_C = p_B(2p_A + p_B + 2p_C) \end{aligned}$$

Since $F_{ST} = 0$, possible coancestry between $X$ and one or both of $Q$ and $K$ is ignored. If $X$, $Q$, and $K$ are all assumed to have a common level of coancestry measured by $F_{ST}$, then we need to take into account the four alleles (AABC) already observed in $Q$ and $K$, so that

$$R_X = 2\text{P}(\text{AB} \mid \text{AABC}) + \text{P}(\text{BB} \mid \text{AABC}) + 2\text{P}(\text{BC} \mid \text{AABC}), \tag{28}$$

where each probability is for an ordered sequence of two alleles. Just as for the derivations of (19) and (20), the conditional probabilities of (28) can be evaluated using the sampling formula, (12). For example $\text{P}(\text{AB} \mid \text{AABC})$ is the probability of observing A then B in two further draws from a population, when a sample of size four has already been observed to be AABC. This probability can be computed as the product of instances of (12) with $m = 2$ and $n = 4$, and with $m = 1$ and $n = 5$. Working similarly for the other two terms we obtain

$$\begin{aligned} R_X &= 2\frac{(2F_{ST} + (1-F_{ST})p_A)(F_{ST} + (1-F_{ST})p_B)}{(1+3F_{ST})(1+4F_{ST})} \\ &\quad + \frac{(2F_{ST} + (1-F_{ST})p_B)(F_{ST} + (1-F_{ST})p_B)}{(1+3F_{ST})(1+4F_{ST})} \\ &\quad + 2\frac{(F_{ST} + (1-F_{ST})p_B)(F_{ST} + (1-F_{ST})p_C)}{(1+3F_{ST})(1+4F_{ST})} \\ &= \frac{(F_{ST} + (1-F_{ST})p_B)(8F_{ST} + (1-F_{ST})(2p_A+p_B+2p_C))}{(1+3F_{ST})(1+4F_{ST})} \end{aligned}$$

All three possibilities for $\mathcal{G}_X$ involve alleles that have already been observed in the genotypes of either $Q$ or $K$. Thus, possible coancestry has the effect of increasing[2] the probability of all possible $\mathcal{G}_X$ over the $F_{ST} = 0$ case.

In the above example, the CSP includes one allele not shared with $K$, and hence which must have come from Q/X. There are essentially only two other cases: that zero and two alleles from Q/X can be determined by subtracting $K$'s alleles from the CSP. An example of the former situation arises when both the CSP = $\mathcal{G}_Q$ = AB. Then $\mathcal{G}_X$ can be any of AA, AB, or BB. If instead, CSP = ABC and $\mathcal{G}_K$ = AA, then $\mathcal{G}_X$ = BC.

### 5.2.5 Two contributors: Q/X and U

Here, difficulties arise because of the number of scenarios to explain the components of the mixture. The relevant LR may be

$$R_X = \frac{\mathrm{P}(\mathrm{CSP}|\mathcal{G}_Q,\ X \text{ and } U \text{ are the sources})}{\mathrm{P}(\mathrm{CSP}|\mathcal{G}_Q,\ Q \text{ and } U \text{ are the sources})}. \tag{29}$$

The numerator requires summation over the conditional genotype probabilities, given $\mathcal{G}_Q$, of all $\mathcal{G}_X$ and $\mathcal{G}_U$ consistent with the CSP. The denominator requires summation over $\mathcal{G}_U$ only. Both summations need to take into account any relatedness among $X$, $Q$ and $U$.

If two co-defendants, $Q1$ and $Q2$, are both alleged to have contributed DNA to the crime sample, then a court may be interested in the strength of evidence for, say, $Q1$ to be a contributor of DNA to the crime sample both with and without assuming that $Q2$ is also a contributor, which could be addressed using the following two LRs:

$$\begin{aligned} R_{X2} &= \frac{\mathrm{P}(\mathrm{CSP}|\mathcal{G}_{Q1}, \mathcal{G}_{Q2},\ X \text{ and } Q2 \text{ are the sources})}{\mathrm{P}(\mathrm{CSP}|\mathcal{G}_{Q1}, \mathcal{G}_{Q2},\ Q1 \text{ and } Q2 \text{ are the sources})} \\ R_{XU} &= \frac{\mathrm{P}(\mathrm{CSP}|\mathcal{G}_{Q1}, \mathcal{G}_{Q2},\ X \text{ and } U \text{ are the sources})}{\mathrm{P}(\mathrm{CSP}|\mathcal{G}_{Q1}, \mathcal{G}_{Q2},\ Q1 \text{ and } U \text{ are the sources})}. \end{aligned}$$

Computing each numerator and denominator of these LRs requires summing over all possible genotypes for whichever of $X$ and $U$ is included. Both observed genotypes $\mathcal{G}_{Q1}$ and $\mathcal{G}_{Q2}$ are potentially informative in both LRs due to possible coancestry with $X$.[3]

The two LRs can differ greatly in value. Intuitively, if $Q2$ is a source of the crime-scene DNA then he can explain many of the observed alleles, thus narrowing the possibilities for $\mathcal{G}_X$. In that case $R_{X2}$ would be larger than $R_{XU}$. However, if both deny being a source of the DNA, a court trying both men jointly may take the view that only $R_{XU}$ can be used, since the presence of DNA from $Q2$ cannot be assumed when assessing the case against $Q1$, and vice-versa. These are decisions for the court; the forensic scientist should try to

[2]The probability could decrease if one of $p_A$, $p_B$, or $p_C$ were much larger than 0·5.

[3]$U$ may also have coancestry with $Q1$ or $Q2$, but the effect on numerator and denominator are similar, so the effect on the LR is typically negligible.

foresee the reasonable possibilities for the pairs of hypotheses that the court may wish to compare.

*Example*

Suppose that, at a particular locus, the following are observed:

$$\mathcal{G}_{Q1} = \text{AB} \qquad \mathcal{G}_{Q2} = \text{CC} \qquad \text{CSP} = \text{ABC}$$

The CSP is consistent with $Q1$ and $Q2$ being the sources of the DNA, and with any other pair of individuals whose genotypes include alleles A, B and C only. Then

$$R_{XU} = \frac{\text{P}(ABC \mid \mathcal{G}_{Q1} = \text{AB}, \mathcal{G}_{Q2} = \text{CC}, X \text{ and } U \text{ are the sources})}{\text{P}(ABC \mid \mathcal{G}_{Q1} = \text{AB}, \mathcal{G}_{Q2} = \text{CC}, Q1 \text{ and } U \text{ are the sources})}. \tag{30}$$

Its evaluation will depend on the relationships among $Q1$, $Q2$, and $X$. Here, for simplicity we will assume that all are unrelated, and initially we will also set $F_{ST} = 0$, so that $\mathcal{G}_{Q2}$ is irrelevant to $R_{XU}$.

The numerator of (30) is $12p_Ap_Bp_C(p_A+p_B+p_C)$. To see this, consider the event that it is the C allele that arises twice in the genotypes of $X$ and $U$: this could result from a CC homozygote and an AB heterozygote (probability $4p_Ap_Bp_C^2$; one factor of two comes from the heterozygote, the other from the orderings of the two genotypes), or from an AC heterozygote and a BC heterozygote (probability $8p_Ap_Bp_C^2$), giving a total probability of $12p_Ap_Bp_C^2$. The expression for the numerator comes from combining this with the two terms corresponding to the A and the B alleles being represented twice.

The denominator of (30) is the probability that $\mathcal{G}_U \in \{\text{AC, BC, CC}\}$, which is $p_C(2p_A+2p_B+p_C)$, and so

$$R_X = \frac{12p_Ap_B(p_A+p_B+p_C)}{2p_A+2p_B+p_C}.$$

If $p_A = p_B = p_C = p$ then $R_X$ reduces to $36p^2/5$, which takes minimum value 1.25 when $p = 1/3$: in this case the evidence is of little value; however, if $p$ is small the evidence is stronger: $R_X = 13.9$ if $p = 0{\cdot}1$, and $R_X = 55.6$ if $p = 0{\cdot}05$.

Now assume that $Q1$, $Q2$, $X$, and $U$ are all unrelated, but are drawn from the same subpopulation for which the level of coancestry, relative to the population allele frequencies, can be characterised by a given value of $F_{ST}$. Now, the probability that it is the C allele that arises twice in the genotypes of $X$ and $U$, given the observed genotypes of $Q1$ and $Q2$, is 12 times the probability that an ordered sample of size four is ABCC, given that a sample of size four has already been observed to be ABCC. Using the sampling formula (12) four times leads to

$$12\frac{(F_{ST}+(1-F_{ST})p_A)(F_{ST}+(1-F_{ST})p_B)(2F_{ST}+(1-F_{ST})p_C)(3F_{ST}+(1-F_{ST})p_C)}{(1+3F_{ST})(1+4F_{ST})(1+5F_{ST})(1+6F_{ST})}$$

and the remaining two terms of the numerator may be computed similarly. The denominator of $R_X$ becomes

$$2\frac{(2F_{ST}+(1-F_{ST})(p_A+p_B))(F_{ST}+(1-F_{ST})p_C)}{(1+3F_{ST})(1+4F_{ST})}+\frac{(2F_{ST}+(1-F_{ST})p_C)(3F_{ST}+(1-F_{ST})p_C)}{(1+3F_{ST})(1+4F_{ST})}.$$

The final result for $R_X$ is tedious to write down but easy to compute. For example, when $p_A = p_B = p_C = 1/3$, the value of 0·8 at $F_{ST} = 0$ increases only slightly to a maximum of 0·8007 at $F_{ST} \approx 1\%$, and then declines as $F_{ST}$ increases. Thus the assumption of a large level of coancestry among $Q1$, $Q2$, and $X$ in this setting *strengthens* the case against $Q1$. However, if $p_A = p_B = p_C = 0{\cdot}1$ then increasing $F_{ST}$ weakens the case against $Q1$: $R_X$ increases from 0·072 at $F_{ST} = 0$ to 0·094 at $F_{ST} = 1\%$ and 0·17 at $F_{ST} = 5\%$.

### 5.2.6 Quantitative interpretation of mixtures

Gill et al. (2006) describe a set of procedures, based on that proposed by Clayton et al. (1998) for interpreting two-contributor mixtures that involves using peak heights from the epg to estimate the heterozygote balance for each pair of alleles. Possible genotype configurations with extreme heterozygote imbalance are then excluded from consideration. For example, for the epg of Figure 4 we noted above in Section 5.2.1 that the peak heights at several loci indicated four distinct alleles, of which two, say alleles A and B, gave strong signals (with approximately equal peak heights), and two, say C and D, gave weak signals (also with approximately equal peak heights). The qualitative approach discussed above would count all three genotype pairs consistent with these alleles (AB,CD; AC,BD; and AD,BC) as equally plausible. The semi-quantitative approach of Clayton et al. (1998) would regard only the AB,CD genotype pair as plausible, because it pairs up the alleles with the two strong and the two weak signals. Congruent genotype combinations across loci are determined through estimating mixture proportion for each contributor.

This semi-quantitative approach seems to be widely-employed in practice, but it is not without difficulties. Estimating mixture proportion relies on an assumption that this proportion is approximately the same across loci. This seems a reasonable assumption prior to the PCR step in the STR typing procedure, and indeed Gill et al. (1998) examined the estimation of mixture contribution proportions and found that consistency across loci is the norm. However, Bill et al. (2005) cite internal FSS data suggesting that single locus mixture proportion estimates can vary by a factor of 0.35 compared to a global estimate, which can result in invalid inferences about plausible genotype pairs at these loci. Further, deciding which genotype configurations are consistent with the mixture proportions is difficult in some borderline cases. For further description and criticisms, see Buckleton et al. (2005).

As a result Gill et al. (1998) went on to suggest a more quantitative approach in which a weight was given to each possible genotype allocation according to its plausibility given the estimated mixture proportion. The ideal, fully-quantitative approach to assessing weight-of-evidence for STR mixtures would involve all the information contained in the epg. Evett et al. (1998) set out a framework for analyses taking peak areas into account. The approach has been expanded on, with software in development or available from multiple authors utilising peak heights (Cowell et al., 2014; Bright et al., 2013; Puch-Solis et al., 2013; Perlin et al., 2011). A consensus on the modelling assumptions for such analysis has yet to be reached, although progress has been made.

## 5.3 LRs for low template DNA profiles

Consider first a single profiling run in which the CSP showed two alleles, $A$ and $B$. The LR (20) comparing $H_Q$ and $H_X$ (page 36) when $\mathcal{G}_Q = AB$ and $F_{ST} = 0$ is

$$\frac{P(\text{CSP} = AB|\mathcal{G}_Q = AB, H_X)}{P(\text{CSP} = AB|\mathcal{G}_Q = AB, H_Q)} = 2p_A p_B.$$

Dropout refers to any allele of a hypothesized contributor that is not observed in the CSP. If CSP $= A$, and low epg peak heights suggest that dropout is possible, then under a standard model (Balding and Buckleton, 2009; Gill et al., 2012) the LR can be written as

$$\frac{P(\text{CSP} = A|\mathcal{G}_Q = AB, H_X)}{P(\text{CSP} = A|\mathcal{G}_Q = AB, H_Q)} = \frac{p_A^2(1-D_2) + 2p_A(1-p_A)D(1-D)}{D(1-D)} \tag{31}$$

where $D$ denotes the probability of dropout for a heterozygote allele, while $D_2$ denotes the probability of a homozygote dropout. The denominator is the probability that the $B$ allele of $Q$ has dropped out ($D$), while the $A$ has not ($1-D$). In the numerator, either $X$ is $AA$ and there has been no dropout (1st term), or (2nd term) $X$ is heterozygous but the non-$A$ allele has dropped out.

Logically, $D$ in the denominator of the LR is different from $D$ in the numerator. However, typically a similar range for $D$ is supported under both hypotheses and they are often taken to be equal for illustrative calculations (Gill et al., 2007). The value of $D$ may vary over alleles because of the effects of degradation, with longer STR alleles having higher $D$ (Tvedebrink et al., 2012).

Dropin refers to an allele in the CSP that is not included in the genotype of any hypothesized contributor. Dropins can be modelled as independent events with probability $C$ (Curran et al., 2005). When both dropout and dropin are possible, all genotypes are consistent with every CSP, and the likelihood under $H_X$ requires summation over every possible genotype for $X$.

# 6 Some legal issues

The weight-of-evidence formula (8) has important implications for the careful reporting of DNA profile evidence. We briefly discuss some of these issues here.

## 6.1 The role of the expert witness

Any individual can make their own assessment of the probability $\text{P}(H_Q|E_d, E_o)$ that the defendant is guilty, based on the DNA evidence $E_d$ and any background information $E_o$ that they feel appropriate. A juror's reasoning is, however, constrained by legal rules. For example, although it may be reasonable to believe that the fact that a person is on trial makes it more likely that they are guilty, a juror is prohibited from reasoning in this way

(to avoid double-counting of evidence). It is therefore usually regarded as inappropriate for an expert witness to report to the court their own assessment of the probability either that the defendant is guilty or, what is often effectively equivalent, that the defendant is the source of the crime profile. A legal rule prohibiting this practice is sometimes called the "ultimate issue rule" (for a discussion see Robertson & Vignaux 1995).

The expert witnesses' primary role is to advise the court on appropriate values for the LRs $R_X$ for various $X$, leaving the jurors to weigh these values together with the other evidence. Although jurors are not required to reason within the logical framework provided by probability theory, it may be helpful for an expert witness to give some explanation of this framework so that the option is fully available to them. The endorsement of any particular values for the $w_X$ should usually be avoided, since this involves the juror's assessment of the non-DNA evidence and is thus usually outside the domain of the expert witness.

Although there are in principle many different LRs, in practice it may suffice to report only a few important values. These might include the values corresponding to a brother of the defendant, another close relative such as a cousin, a person apparently unrelated to the defendant but with a very similar ethnic background and a person completely unrelated to the defendant. Methods for calculating LRs were discussed above in Section 4.

## 6.2 The prosecutor's fallacy

The prosecutor's fallacy is a logical error that can arise when reasoning about DNA profile evidence. You may be aware of the error in elementary logic of confusing "A implies B" with "B implies A". For example, if A denotes "is a cow" and B denotes "has 4 legs", then (ignoring rare anomalies) A implies B, but the converse doesn't hold. The prosecutor's fallacy is similar to this logical error, but is in terms of probabilities.

The fallacy consists of confusing P(A|B) with P(B|A). If it is accepted that the probability that the criminal is very tall is 90%, it doesn't follow that the next very tall man you meet has 90% probability of being the criminal. The correct way of obtaining P(A|B) from P(B|A) is given by the appropriate version of Bayes Theorem, for example the weight-of-evidence formula (8).

Transcripts of actual court cases have in the past very often recorded statements that indicate that the match probability is being confused with the probability that the defendant is innocent (see box below). This error, which can be extremely detrimental to defendants, has led to successful appeals in the UK.

## 6.3 The defendant's fallacy

Another error of logic that can arise in connection with DNA evidence usually favours the defendant and is consequently dubbed the "defendant's fallacy". Suppose that a crime occurs in a nation of 100 million people and a profile frequency is reported as 1 in 1 million. The fallacy consists of arguing that, since the expected number of people in the

## The prosecutor's fallacy: examples

In the quotations below, an expert witness has made a statement about the probability that the defendant is not the source of the crime profile, which is outside the domain of an expert witness.

- "I can estimate the chances of this semen having come from a man other than the provider of the blood sample ... less than 1 in 27 million".
- "The FBI concluded that there was a 1 in 2,600 probability that the semen ... came from somebody other than Martinez".

nation with a matching profile is 100, the probability that the defendant is guilty is at most only 1 in 100, or 1%.

This conclusion would be valid only if, ignoring the DNA evidence, every person in the nation is equally likely to be the culprit. In the notation of the weight-of-evidence formula, each $w_X$ is equal to one. In practice, such an assumption is rarely reasonable. Even if there is little or no directly incriminating evidence beyond the DNA profile match, there is always background information presented in evidence, such as the location and nature of the alleged offence, that will make some individuals more plausible suspects than others.

A closely related fallacy consists of arguing that, since it is expected that many people in the nation share the profile, the DNA evidence is almost worthless and should be ignored. Correct use of the DNA evidence formula avoids these, and other, fallacies.

# References


Balding, D. 1999. When can a DNA profile be regarded as unique? *Sci Justice*, 39:257–260.

Balding, D. and Buckleton, J. 2009. Interpreting low template DNA profiles. *Forens. Sci. Intern.: Genet.*, 4:1–10.

Bill, M., Gill, P., Curran, J., Clayton, T., Pinchin, R., Healy, M., and Buckleton, J. 2005. Pendulum—a guideline-based approach to the interpretation of str mixtures. *Forensic science international*, 148(2):181–189.

Bright, J.-A., Taylor, D., Curran, J. M., and Buckleton, J. S. 2013. Developing allelic

and stutter peak height models for a continuous method of dna interpretation. *Forensic Science International: Genetics*, 7(2):296–304.

Buckleton, J., Triggs, C., and Walsh, S. 2005. *Forensic DNA Evidence Interpretation*. CRC Press.

Clayton, T., Whitaker, J., Sparkes, R., and Gill, P. 1998. Analysis and interpretation of mixed forensic stains using dna str profiling. *Forensic Science International*, 91(1):55–70.

Cowell, R., Graversen, T., Lauritzen, S., and Mortera, J. 2014. Analysis of forensic dna mixtures with artefacts. *Analysis*, 5:6.

Curran, J., Gill, P., and Bill, M. 2005. Interpretation of repeat measurement DNA evidence allowing for multiple contributors and population substructure. *Forens. Sci. Intern.*, 148(1):47–53.

Curran, J. M., Triggs, C. M., Buckleton, J., and Weir, B. 1999. Interpreting dna mixtures in structured populations. *Journal of Forensic Sciences*, 44:987–995.

Evett, I., Buffery, C., Wilcott, G., and Stoney, D. 1991. A guide to interpreting single locus profiles of DNA mixtures in forensic cases. *J. Forens. Sci. Soc.*, 31:41–47.

Evett, I., Gill, P., and Lambert, J. 1998. Taking account of peak areas when interpreting mixed dna profiles. *J. Forens. Sci.*, 43:62–69.

Fung, W. K. and Hu, Y.-Q. 2000. Interpreting forensic dna mixtures: allowing for uncertainty in population substructure and dependence. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 163(2):241–254.

Fung, W. K. and Hu, Y.-Q. 2002. The statistical evaluation of dna mixtures with contributors from different ethnic groups. *International journal of legal medicine*, 116(2):79–86.

Fung, W. K. and Hu, Y.-Q. 2004. Interpreting dna mixtures with related contributors in subdivided populations. *Scandinavian journal of statistics*, 31(1):115–130.

Gill, P., Brenner, C., Buckleton, J., Carracedo, A., Krawczak, M., Mayr, W., Morling, N., Prinz, M., Schneider, P., and Weir, B. 2006. DNA Commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. *Forens. Sci. Intern.*, 160(2-3):90–101.

Gill, P., Gusmão, L., Haned, H., Mayr, W., Morling, N., Parson, W., Prieto, L., Prinz, M., Schneider, H., Schneider, P., and Weir, B. 2012. DNA commission of the International Society of Forensic Genetics: Recommendations on the evaluation of STR typing results that may include drop-out and/or drop-in using probabilistic methods. *Forens. Sci. Intern.: Genet.*, 6(6):679–88.

Gill, P., Kirkham, A., and Curran, J. 2007. LoComatioN: A software tool for the analysis of low copy number DNA profiles. *Forens. Sci. Intern.*, 166:128–138.

Gill, P., Sparkes, R., Pinchin, R., Clayton, T., Whitaker, J., and Buckleton, J. 1998. Interpreting simple STR mixtures using allele peak areas. *Forens. Sci. Intern.*, 91(1):41–53.

Haned, H., Pène, L., Lobry, J. R., Dufour, A. B., and Pontier, D. 2011. Estimating the number of contributors to forensic dna mixtures: does maximum likelihood perform better than maximum allele count? *Journal of forensic sciences*, 56(1):23–28.

Hu, Y.-Q. and Fung, W. K. 2003. Evaluating forensic dna mixtures with contributors of different structured ethnic origins: a computer software. *International journal of legal medicine*, 117(4):248–249.

Mortera, J., Dawid, A. P., and Lauritzen, S. L. 2003. Probabilistic expert systems for dna mixture profiling. *Theoretical population biology*, 63(3):191–205.

Perez, J., Mitchell, A. A., Ducasse, N., Tamariz, J., and Caragine, T. 2011. Estimating the number of contributors to two-, three-, and four-person mixtures containing dna in high template and low template amounts. *Croatian medical journal*, 52(3):314–326.

Perlin, M. W., Legler, M. M., Spencer, C. E., Smith, J. L., Allan, W. P., Belrose, J. L., and Duceman, B. W. 2011. Validating trueallele® dna mixture interpretation*,. *Journal of forensic sciences*, 56(6):1430–1447.

Puch-Solis, R., Rodgers, L., Mazumder, A., Pope, S., Evett, I., Curran, J., and Balding, D. 2013. Evaluating forensic dna profiles using peak heights, allowing for multiple donors, allelic dropout and stutters. *Forensic Science International: Genetics*, 7(5):555–563.

Torres, Y., Flores, I., Prieto, V., López-Soto, M., Farfán, M. J., Carracedo, A., and Sanz, P. 2003. Dna mixtures in forensic casework: a 4-year retrospective study. *Forensic science international*, 134(2):180–186.

Tvedebrink, T., Eriksen, P., Mogensen, H., and Morling, N. 2012. Statistical model for degraded DNA samples and adjusted probabilities for allelic drop-out. *Forens. Sci. Intern.: Genet.*, 6:97–101.