Comment: Some Causes for Concern about DNA Profiles

David J. Balding, Peter Donnelly and Richard A. Nichols

1. INTRODUCTION

The author has surveyed many aspects of the current debate on the forensic use of DNA profiles. It is our view that, despite the profuse literature on the topic, several important issues have not yet been adequately addressed. We hope that the present paper will assist in widening the debate within the statistical community and hence lead to clarification of these issues.

2. INTERPRETATION OF THE LIKELIHOOD RATIO

Only recently has a consensus emerged that the appropriate measure of the strength of forensic evidence is the likelihood ratio. However, some of the implications have not been fully appreciated. How is the likelihood ratio to be interpreted? What is to be made of the various likelihood ratios which the author reports, each contrasting H_1 with a different H_0? How should the effect of other, possibly exculpatory, evidence be incorporated with the DNA evidence?

There may have been a tacit assumption by some commentators (e.g., the NRC and Collins et al., 1994) that reported likelihood ratios should be the basis of hypothesis tests. We believe that such an approach is inappropriate. The most serious concern is that it is extremely difficult in such an approach to allow for the effect of the non-DNA evidence. Consider two hypothetical cases of assault. In one case the assailant is recognized by the victim to be a man living at a neighbouring address. The man is duly arrested and his profile is found to match that of the crime sample. In the second case the victim did not see the assailant. A DNA profile match is discovered "by chance," through a forensic scientist noticing a similar profile from a man living in another part of the country. However, this man produced an apparently valid alibi and subsequent investigation could reveal no link with the crime. How could these very different sets of evidence be accounted for in a hypothesis testing framework, if the relevant likelihood ratios were the same in each case?

We believe that the only logical method of weighing the DNA evidence in conjunction with the other evidence is to use Bayes' rule. If one accepts that the concept of probability can be applied to hypotheses such as H_0 and H_1, as for example the author does in Section 2.1, then it is simply a matter of elementary probability that Bayes' rule gives the correct method for updating these probabilities in the light of the DNA evidence. The controversy over Bayesian techniques in other areas of statistics is not directly relevant here.

In the legal context, Bayes' rule makes clear the distinction between the domain of the expert witness, the likelihood ratio, and the domain of the court, the assessment of other evidence. There is a substantive debate about the extent to which probabilities and mathematical reasoning, in any form, are appropriate in court (e.g., Tribe, 1971) and the extent to which juries should be educated in—and encouraged to use—Bayes' rule. We do not address this debate here. Our point is that concerns which arise in consequence of the logical analysis are legitimate regardless of the method of analysis actually adopted by juries. Nevertheless there are real dangers in reporting a likelihood ratio to an untrained jury without an explanation of its interpretation (Kaye and Koehler, 1991; Donnelly and Balding, 1994).

One crucial consequence of the appropriate interpretation of the likelihood ratio concerns the widely held view, expressed by the author in Section 8, that an error of one or two orders of magnitude "will have little practical impact on likelihood ratios as large as several million." This is incorrect. The assessment of the strength of the other evidence lies in the domain of the court, not that of the expert witness. It is quite plausible that in some cases this would correspond to extremely small prior odds. For example, in a case in which there is little or no evidence other than the DNA match, it may be reasonable to estimate that there are 10,000 individuals who, before examining the DNA profiles, were just as likely as the

David J. Balding is Lecturer in Probability and Statistics, Peter Donnelly is Professor of Mathematical Statistics and Operational Research and Richard A. Nichols is Lecturer in Genetics at Queen Mary & Westfield College, University of London, Mile End Road, London E1 4NS, United Kingdom.

defendant to be the source of the crime sample. Similarly, if there is convincing evidence which tends to exonerate the defendant, the appropriate prior odds might be 1 in 10,000. In such cases a likelihood ratio of 100,000 may well lead to acquittal while one of 10 million may suffice for a conviction. The apparently "small" difference between these likelihood ratios could be critical.

3. CORRELATIONS

Another essential feature of inference for forensic identification is that the relevant probabilities are *conditional* probabilities. The probability that a "random" person has a particular DNA profile is not directly relevant. Instead, under general assumptions, Balding and Donnelly (1994) show that the posterior odds on the hypothesis of identity H_1 are given by

$$(1) \qquad \frac{1}{\text{Odds}(H_1)} = \sum_{i \neq s} P(\mathcal{G}_i \mid \mathcal{G}_s) \frac{P(C = i)}{P(C = s)},$$

in which we introduce $\mathcal{G}_i$ for the event that the ith person has the profile and s and C for the label (or name) of, respectively, the defendant and the person who was the source of the crime sample. For convenience, we refer to this latter individual as the culprit but note that this is not necessarily the case. The summation in (1) is over every possible culprit. Each probability in (1) is conditional on all the other information, but we suppress the conditioning in the notation. For example, $P(C = s)$ denotes the probability that the defendant is the culprit based on the other information except the defendant's DNA profile.

The point of equation (1) is that one needs to consider each possible culprit and to assess the probability that they would match the crime profile, given that the defendant does. These *conditional* probabilities should then be weighted in a way which depends on the probabilities, given the other evidence, that that individual is the culprit.

Statistical analyses of forensic databases are not directly relevant to an assessment of the conditional probabilities $P(\mathcal{G}_i \mid \mathcal{G}_s)$. It is thus necessary for the statistician to model these conditional probabilities. In the forensic context, particularly for genetic traits, positive correlations arise which can cause the appropriate conditional probability to be substantially larger than the unconditional value.

The strongest positive correlations in (1) arise for those individuals i who are closely related to the defendant s. Most current analyses effectively ignore these terms. This can be very detrimental to the defendant, *even in cases when there is no specific evidence to cast suspicion on his relatives*. Certain close relatives may in some cases be excluded by other evidence. However, in most cases relatives of the defendant are no less likely to be guilty *a priori* than a "random" unrelated person, and often the circumstances of the case are such that they are more likely.

This effect is most important for siblings. When i is a sibling of s, the value of $P(\mathcal{G}_i \mid \mathcal{G}_s)$ is substantially larger than the corresponding probability for unrelated individuals. For example, in a recent case (*HMA v. Aslam,* 1993), the forensic scientist reported a match probability for unrelated individuals, based on three single-locus probes, of 1 in 49,000. At trial, he accepted that the probability that a particular brother of the defendant would match was about 1 in 16. As it happened, the defendant had five brothers. In other cases there may be large numbers of other relatives, such as cousins or half-siblings.

Even if a large number of individuals, including the defendant's relatives, are considered, based on the other evidence, equally likely to be the culprit, then often the calculation of the posterior odds (1) will be dominated by the effect of the relatives (Donnelly, 1992; Balding and Nichols, 1994). An innocent defendant will often be reluctant for good reason to raise the possibility that his relatives are guilty. However, the effects of this possibility are so important that, in the interests of justice, the prosecution should make reasonable allowance for it. Unfortunately, this is rarely the case in current practice.

After relatives, the conditional probabilities $P(\mathcal{G}_i \mid \mathcal{G}_s)$ will be largest for individuals i who share ancestors on an evolutionary time scale. The evidence which pointed to the defendant will often suggest that if the defendant were innocent, then the culprit would be similar in some respects and it was this similarity which led to the false accusation. For example, the culprit may well have a similar physical description and/or live in the same neighbourhood and/or frequent the same institutions as the innocent defendant and so forth. Thus the terms in (1) corresponding to i and s belonging to the same sub-population can make a substantial contribution to the sum since both $P(\mathcal{G}_i \mid \mathcal{G}_s)$ and $P(C = i)$ will be relatively large. We thus disagree with the author's claim (Section 4) that the ethnicity of the defendant is irrelevant to forensic inference. Assumption 2 of Section 4.1 is, in general, both false and detrimental to the defence (Balding and Donnelly, 1994). Further, it may often be appropriate, and will usually be cautious, to replace the $P(\mathcal{G}_i \mid \mathcal{G}_s)$ for nonrelatives in the sum (1) by the value appropriate for i and s belonging to the same subpopulation.

4. ASSESSING GENETIC DIFFERENTIATION

It is not necessary to believe in "gross violations of the assumption of independence," "radically dif-

ferent profile probabilities" and "extreme population heterogeneity" in order to cast doubt on the author's conclusion that "the tremendous genetic variability among individuals obviates concerns arising from minor violations of modeling assumptions." In order to evaluate (1) it is necessary to address the question of genetic differentiation at a range of levels of stratification, including in some cases small, partly isolated subpopulations. Published data, [e.g., Figure 3 of Balazs et al. (1989), Figure 2 of Buffery et al. (1991) and the author's Figure 4b] show that at some loci, there is a large interracial differentiation. This observation raises the question of how much genetic variation there might be at finer scales of population subdivision.

Answering this question will require substantial surveys at appropriate levels of stratification at each of the loci in forensic use. The population genetic analyses discussed by the author are mostly restricted to broad levels of population stratification and there are very few studies available at the finer levels of stratification which will usually be relevant to forensic inference. One such study is that of Krane et al. (1992), which suggested a level of differentiation within the Caucasian classification which is sufficient to cause concern about current forensic practice. We are aware of criticisms of the study and hence eagerly await the results of further surveys. However, in the interim, only cautious assumptions should be made about levels of genetic differentiation. Unfortunately, the assumptions underlying much of current practice are not cautious.

There are currently two methods established in courts which attempt to allow for genetic differentiation. As the author notes, the so-called ceiling principle advocated by the NRC report, has been widely criticized. The other method, essentially the author's equation (12), was proposed for forensic applications by two of us (Nichols and Balding, 1991). It has been adopted in some U.K. cases and has had some success, in the sense of acceptance by both prosecution and defence. It models the correlations due to common ancestry in terms of the parameter θ_S, similar to the population geneticist's F_{ST}. Appropriate values of θ_S may well be locus-specific, due to differences in the mutation mechanisms, but values will be positively correlated across loci.

The author presents evidence that θ_S is small, citing a point estimate of 0.15% in a Caucasian database. Such indirect estimates of θ_S, which in effect measure excess homozygosity, may depend sensitively on assumptions about apparent homozygotes. Further, it is important to recall that forensic databases are not random samples. In any case, analyses of large heterogeneous databases cannot lead to values of θ_S appropriate for differing levels of population stratification. Studies of individuals known to belong to distinct ethnic groups are required. Such studies have been conducted at traditional loci, such as the distribution of the frequency of blood groups in British regions defined on a fine scale (Kopec, 1970). A point estimate of θ_S obtained from this data is also low, about 0.3%. However, the 56 samples from the Glasgow region (a major urban conurbation) vary much more substantially from national values, presumably due to Glasgow's unique history of migration from Ireland and the Scottish Highlands. Many cities in other countries also have such distinctive histories. Further, genetic differentiation is known be more marked in African and Asian populations than among Caucasians (Cavalli-Sforza and Piazza, 1993).

Even when a point estimate of θ_S is modest, the effect on whole-profile match probabilities can be important. One reason is that θ_S is not a constant but varies from one subpopulation to another according to its evolutionary history. Thus evaluation of $P(\mathcal{G}_i \mid \mathcal{G}_s)$ requires integration over the distribution representing the possible values of θ_S. As high powers of θ_S are involved, values in the tail of the distribution will have a disproportionate impact (Balding and Nichols, 1994).

We believe that the author's equation (19), dealing with the effect of relatives, is inappropriate since it ignores an important effect of the conditioning. Instead, an equation analogous to (12) is particularly appropriate here. This is because marriage partners are chosen preferentially within the same subpopulation, and hence the defendant's profile can impart information about relative frequencies in the relevant sub-population. For details of the extension of (19) to allow for this effect, see Balding and Nichols (1994).

We agree with Evett, Scranage and Pinchin (1993) that attempts to establish that genetic differentiation effects are unimportant by failing to reject hypotheses of independence are of only limited usefulness. Acceptance of such null hypotheses is somewhat reassuring, but the tests will have negligible power against many alternatives that are important in the forensic context. The null hypothesis in such tests is indubitably false. The question of interest concerns the magnitude of the difference between $P(\mathcal{G}_i \mid \mathcal{G}_s)$ and $P(\mathcal{G}_i)$ and hypothesis tests bear only indirectly on this question.

5. OTHER ISSUES

5.1 Uncertainty

Another source of positive correlations which is frequently overlooked is uncertainty about the relative frequency of the profile. Most current analyses ignore many sources of uncertainty and this is detri-

mental to the defendant (Balding and Donnelly 1994, Balding and Nichols, 1994).

There may be large classes of individuals for whom it is reasonable to assume that possession of the profile is exchangeable. In this case De Finetti's theorem can be useful in modelling profile possession. It is natural to interpret the directing measure in the associated De Finetti representation as summarizing the uncertainty about the probability of profile possession. An important consequence is that uncertainty induces positive correlations which decrease the posterior odds on H_1.

There are several sources of uncertainty in the forensic use of DNA profiles. There is uncertainty due to sampling effects and the fact that forensic databases are not random samples. Profile relative frequencies are estimated in terms of the product of sample allele frequencies and there is uncertainty about the validity of the independence assumption underlying this estimation procedure. Perhaps the most important source of uncertainty is the fact that the population from which the database is drawn may not be representative of the individuals who contribute most to the sum in (1).

5.2 Databases

The author discusses in Section 7.1 the question of the appropriate analysis in the case that the defendant has been identified by searching a DNA profile database. The reasoning which led to (1) can be extended to incorporate this situation (Balding and Donnelly, 1994). An analysis shows that, loosely speaking, the weight of the evidence against the defendant is slightly increased because of the individuals who have been excluded from suspicion by the search. Note, however, that there may be little or no corroborating evidence in the case that the defendant is identified through a database search.

5.3 Alternative Analyses

Although in principle an evaluation of the likelihood ratio based directly on the data and avoiding the declaration of "match" is attractive, the advantages of this approach may well be outweighed by the additional difficulties it introduces, for example, the additional technical complexity. The author notes the fundamental problem with the BEP method. Our own simulations indicate that it can lead to substantial overstatements of the likelihood ratio. However, the DRR method also has problems: estimation of the genotype distribution is an ill-posed inverse problem and maximum likelihood estimates can be unstable. Variability in the flanking region and within the repeat unit may cast doubt on the assumption that the support of the genotype distribution is restricted to integer multiples of a fixed repeat unit.

There are good genealogical reasons to believe that the underlying distribution is discrete. The importance of this effect depends on the mutation mechanisms which are not yet well understood.

5.4 Uniqueness

It is hard to imagine how any DNA profiles could be claimed to be "unique" without a survey of all the world's profiles. We note that dermal fingerprints have an advantage over DNA profiles in that a fingerprint is not entirely inherited and hence correlations due to relatedness and shared ancestry are less problematic. Because of these correlations, however, the DNA profiles in current forensic use are not unique.

The discussion of the question of uniqueness in numerous publications which followed the Collins case is flawed (Balding and Donnelly, 1994). In particular, what the author refers to as the "appropriate calculation" has been established to be misleading (Dawid, 1994). We believe that pursuit of the question of uniqueness is unlikely to be fruitful. The appropriate question concerns only how much the evidence changes the probability of hypothesis H_1.

ACKNOWLEDGMENTS

This work supported in part by the U.K. Science and Engineering Research Council under Grants GR/F 98727, GR/G 11101 and B/AF/1255.