Lindsey A. Foreman,[1] *Ph.D., Adrian F. M. Smith,*[1] *Ph.D., and Ian W. Evett,*[2] *D.Sc.*

Bayesian Validation of a Quadruplex STR Profiling System for Identification Purposes

REFERENCE: Foreman LA, Smith AFM, Evett IW. Bayesian validation of a quadruplex STR profiling system for identification purposes. J Forensic Sci 1999;44(3):478–486.

ABSTRACT: A quadruplex system for determining the genetic profile of an individual at four short tandem repeat (STR) loci has recently been introduced into forensic casework by the Forensic Science Service (FSS), primarily for the purposes of forensic identification. Data have been collected under this system from the three racial groups of most relevance in casework in the UK: Caucasian, Afro-Caribbean and Asian (from the Indian subcontinent). These data are utilized in calculations to quantify the evidential strength of a DNA match between suspect and crime scene sample, say, through the evaluation of a likelihood ratio (LR). Previous papers (1,2) have studied the databases via classical statistical methods. However, we focus on a Bayesian approach (3) to validation of the data for LR evaluation in two main cases: when individuals being compared are either (i) completely unrelated, or (ii) members of the same racial group subpopulation. Empirical studies are conducted to establish the robustness of proposed models and obtain efficient and adequate approximations to the LR calculations. This involves the use of statistical simulation methods to determine the suitability of the product rule and Bayesian inference for coancestry coefficients in the absence of subpopulation data.

KEYWORDS: forensic science, STR markers, Bayesian statistics, likelihood ratio, product rule, population heterogeneity

When DNA evidence is presented in court, it is usually quantified in terms of a match probability or likelihood ratio (LR). There has been much discussion in the literature on the appropriate method of evaluating these quantities (4,5) and this will typically involve the use of a measure, θ, representing the level of coancestry within a population. The recent NRC (National Research Council) report (4), which provides U.S. guidelines for the evaluation of forensic DNA evidence, suggests adopting a value of $\theta = 0.03$ for all racial groups in conjunction with newly-introduced PCR-based systems (Recommendation 4.1). In this paper we aim to explore the validity of the NRC recommendation for UK data collected from the three main racial groups by the Forensic Science Service (FSS) under a quadruplex (4-locus) STR profiling system (6). These particular data have been analyzed elsewhere using conventional statistical methods (1,2). We focus here on a Bayesian approach to the problem, based on ideas developed in an earlier paper (3).

We use $\mathbf{x}$ as generic notation for the STR profile of an individual typed at M loci. In this paper, we restrict attention to applications of forensic identification; i.e., what evidence exists to suggest that a suspect (S) standing trial is the offender (O) in a criminal incident? If we let $\mathbf{x}^s$ denote the STR profile of the suspect and $\mathbf{x}^o$ denote the STR profile of the offender, then a match occurs if these are equal since there are no complicating factors such as measurement error and coalescence with STRs. The strength of evidence in support of the prosecution hypothesis that offender and suspect are the same person versus the alternative that they are different individuals is then represented by the likelihood ratio:

$$\frac{p(\mathbf{x}^o, \mathbf{x}^s | O = S)}{p(\mathbf{x}^o, \mathbf{x}^s | O \neq S)}$$

If the profiles $\mathbf{x}^o$ and $\mathbf{x}^s$ do not match, we obtain a LR of 0. Conversely, if we observe matching profiles, the form of the LR reduces to

$$LR(\mathbf{x}) = \frac{1}{p(\mathbf{x}^o = \mathbf{x} | \mathbf{x}^s = \mathbf{x}, O \neq S)}$$

The denominator is referred to as the conditional match probability of $\mathbf{x}$ and this relates to the case where offender and suspect are different individuals. Clearly, this probability will be influenced by how closely these two individuals are related; e.g., at one extreme, the match probability will be relatively high when comparing two brothers since they are likely to share very similar genetic characteristics (7) and, in this case, the evidence provided by a match is not as persuasive as when offender and suspect are completely unrelated. We focus on validation of the quadruplex data for computing LRs in two important cases: (i) offender and suspect are completely unrelated, and (ii) offender and suspect are members of the same inbreeding racial group subpopulation.

In the next section we address case (i), where the LR may be evaluated via the well-known *product rule*, provided checks are made on the suitability of the underlying independence assumptions. One possible robustness check is detailed which investigates how LR distributions are affected. A Bayesian approach to LR calculations in case (ii) is then provided and this is based on making inference about coancestry coefficients when data relating to appropriate identified subpopulations is unavailable. Finally, results are presented from implementation of these methods for the FSS quadruplex data, including suggestions for the simplification of techniques which are suitably robust in a forensic setting.

Comparison of Completely Unrelated Individuals

Let us assume the offender belongs to a particular racial group $\mathscr{P}$. If the suspect and offender are taken to be distinct and genetically unrelated either within $\mathscr{P}$ or as members of different racial groups then this may be considered equivalent to assuming their

[1] Department of Mathematics, Imperial College of Science, Technology and Medicine, London, UK.
[2] The Forensic Science Service, London, UK.
Received 5 Feb. 1998; and in revised form 30 June 1998; accepted 14 Sept. 1998.

profiles are independent of each other. Thus, the match probability in this case simply corresponds to the proportion of the profile **x** in $\mathscr{P}$; i.e., $p(\mathbf{x}^o = \mathbf{x}|\mathbf{x}^s = \mathbf{x}, O \neq S) = p(\mathbf{x}^o = \mathbf{x})$. It is then standard practice to assume that Hardy-Weinberg and linkage equilibrium conditions are satisfied so that alleles are inherited independently both within and across loci. The product rule may then be invoked to evaluate complete profile probabilities by essentially multiplying together component genotype probabilities. At a single locus, these are evaluated as

$$\begin{cases} \gamma_j^2 & \text{, for the homozygote } (A_j, A_j) \\ 2\gamma_j\gamma_k & \text{, for the heterozygote } (A_j, A_k) \end{cases} \quad (1)$$

where γ_j and γ_k denote the proportions of alleles A_j and A_k, respectively, as exhibited in the racial group $\mathscr{P}$. Terms such as (1) are then multiplied across all M loci. Henceforth, we shall refer to this as the *independence model*. In real human populations, of course, the idealized assumptions necessary for Hardy-Weinberg and linkage equilibrium to hold exactly never exist and the essential forensic issue is to determine whether (incorrect) use of the independence model could give results which might mislead a court. This is addressed in the following sections.

LR Evaluation

Let $\mathscr{D} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ denote the dataset of STR profiles from n individuals drawn from the racial group of interest, $\mathscr{P}$. Even though such databases are often compiled using convenience samples of individuals, it is usually argued that the underlying genetic samples are effectively random (see Chapter 5 of (4)). Assuming that the within- and between-locus allele independence assumptions hold, our earlier paper (3) describes two methods for evaluating LRs: (a) plug-in estimates, and (b) full Bayesian analysis.

Under the plug-in estimate approach (a), the proportion of each allele is estimated by its relative frequency in the observed dataset $\mathscr{D}$, where the suspect's profile $\mathbf{x}^s = \mathbf{x}$ is temporarily added to the appropriate racial group database as an extra piece of relevant data which has been observed. The relative frequency estimates, $\hat{\gamma}$, are then substituted in (1) and multiplied across loci to obtain the probability of the profile **x** and, hence, by inversion, the corresponding LR.

In general, the Bayesian approach represents our uncertainty about an unknown quantity (the set of allele distributions at each locus, γ, say) through an entire probability distribution. Before observing any data, we may have certain beliefs about the specification of γ based on other studies, intuition, etc. These are represented by the *prior* distribution, $p(\gamma)$. The *likelihood*, $p(\mathscr{D}|\gamma)$, is a function which then explains how the data behave described in terms of γ. In our case, this will be in the form of a product of profile probabilities, one for each individual observed in the dataset $\mathscr{D}$, including the suspect if (s)he is a member of $\mathscr{P}$. After observing the data, we update our beliefs about γ via *Bayes' theorem* and obtain the *posterior* distribution, $p(\gamma|\mathscr{D})$:

$$p(\gamma|\mathscr{D}) \propto p(\gamma) \times p(\mathscr{D}|\gamma) \quad (2)$$

This represents a weighted combination of the data (through the likelihood) and our prior beliefs. If we adopt independent uniform priors for the allele distributions at each locus in γ, this can be interpreted as representing prior ignorance. The posterior distributions resulting from expression (2) can then be shown to be independent and of a standard form (i.e., the so-called Dirichlet distribution). Thus, if m denotes the number of distinct alleles exhibited at a particular locus, the corresponding posterior distribution has the form:

$$\prod_{k=1}^{m} \gamma_k^{\eta_k}$$

where η_k denotes the number of times the allele A_k is observed in the dataset $\mathscr{D}$, after inclusion of the suspect's profile in his racial group database. Therefore, under the Bayesian approach (b), the posterior probability of profile **x** may be found by *integrating* expression (1) with respect to the posterior of the allele distribution at each locus and then multiplying across loci. The term corresponding to a single-locus genotype is given by

$$\begin{cases} \dfrac{(1+\eta_j)(2+\eta_j)}{(m+2n)(1+m+2n)} & \text{, for the homozygote } (A_j, A_j) \\ 2\,\dfrac{(1+\eta_j)(1+\eta_k)}{(m+2n)(1+m+2n)}, & \text{for the heterozygote } (A_j, A_k) \end{cases} \quad (3)$$

if the suspect does not belong to the racial group $\mathscr{P}$ under consideration, and the same expression (3) with n replaced by $n + 1$, if $\mathscr{P}$ *does* correspond to the suspect's racial group. The Bayesian estimate, using (3), of the probability of profile **x** in $\mathscr{P}$ can be shown algebraically and empirically to approximate the plug-in estimate with the suspect's profile being added twice to the database corresponding to his racial group and once to the remaining racial group databases (sometimes referred to as the sampling/size-bias correction). For example, if $\mathscr{P}$ corresponds to a racial group different from the suspect's, the heterozygote expression (3) can be approximated by the plug-in estimate with the suspect's profile temporarily added once to the corresponding database; i.e.

$$2\,\frac{(1+\eta_j)}{(2n+2)}\frac{(1+\eta_k)}{(2n+2)}$$

since typically $n >> m$.

A Simple Test to Check the Robustness of LR Distributions to Allele Independence Assumptions

The independence model involves the multiplication of (1) or (3) across loci and we must show that, within the three racial groups of interest, the assumption of within- and between-locus independence of alleles does not result in misleading LR calculations. Classical statistical significance tests that check adherence to this null "independence" hypothesis are commonly applied; in particular, the exact test (8). However, the idealized concepts of Hardy-Weinberg and linkage equilibrium and, hence, the independence assumptions are never realized in real populations, due principally to the presence of *substructure* caused by preferred mating of individuals within subpopulations. In forensic applications, we are more interested in the practical consequences of adopting (1) and multiplying across loci as a simplified model for evaluating profile probabilities and judge the adequacy of this model based on two different criteria. The first, which we study in the remainder of this section, looks at the effect in terms of an aggregate LR measure which has been employed as an illustrative tool in court to explain the discriminatory power of reported DNA matches. The second criterion is addressed later and investigates the effect on individual LRs of adopting what may be considered a more realistic model for LR evaluation.

Our aim in this section is to compare the distribution of STR profiles in the racial group of interest $\mathscr{P}$ with those seen in "equivalent"

idealized populations, artificially constructed so that alleles are truly independent within and between loci. This may be achieved via comparison of the profile distributions observed in datasets representing such populations as these provide estimates of the underlying population profile distributions. Since the STR profile distribution is discrete and defined over a very large number of possible profiles, it is more feasible to make comparisons by means of the corresponding distribution of LRs (which are the quantities of forensic interest) for profiles in each dataset where these are evaluated using the independence model (1). We extend and develop an idea here which has been applied previously (2,9).

The distribution of LRs exhibited by profiles in the observed database $\mathscr{D}$ computed using the independence model via (1) may be obtained and compared with those distributions typically exhibited under equivalent databases which have been constructed to satisfy the independence assumptions exactly. Such LR distributions may be constructed under two different scenarios: (i) suspect and offender are the *same* individual, and (ii) suspect and offender are *different* individuals. For any database of profiles from *n* individuals, we can simulate cases where suspect and offender are the same person by looking at all "within-person" comparisons. We have *n* such cases and, by computing the LR corresponding to a matching profile equal to each of these *n* individual profiles, the associated LR distribution exhibited by the database may be obtained. Similarly, cases where suspect and offender are different persons may be simulated by performing all $n(n-1)/2$ pairwise "between-person" comparisons of individuals represented in the database. A match is declared when both individuals being compared have identical profiles, in which case the LR associated with the shared profile is recorded. When the two profiles do not match, a LR of 0 is recorded. Under both these scenarios, curves may be constructed that describe the underlying LR distribution exhibited by the database profiles by plotting "the proportion of cases with $LR > \lambda$" vs. "λ", for various values of λ, in an analogous way to cumulative probability distributions. These plots are sometimes referred to as *Tippett diagrams* (10).

The LR curves may be plotted based on the observed dataset of *n* profiles and, similarly, for equivalent databases of size *n* which are artificially generated to ensure they satisfy the independence assumptions exactly. The independence model computes LRs correctly for profiles in such databases. Under the plug-in estimate approach (a), each new database of *n* profiles, $\mathscr{D}^*$, is constructed by independently generating alleles at each locus according to the estimated allele proportions in $\hat{\gamma}$. In this sense, $\mathscr{D}^*$ is "equivalent" to the observed database, $\mathscr{D}$, since it is of the same size and originates from the same set of allele distributions. Under the Bayesian approach (b), we generate a particular set of allele distributions, γ^*, from the posterior distribution $p(\gamma|\mathscr{D})$ and construct the new database $\mathscr{D}^*$ conditional on this set of allele distributions, as for approach (a). This procedure may be repeated *T* times to obtain a representative sample of LR curves typical of those we would expect to see in population datasets satisfying the allele independence assumptions exactly. In particular, the 5th and 95th percentile LR curves may be plotted and compared with the "observed" LR curves.

The purpose of this exercise is simply to construct a typical set of population databases satisfying the independence model exactly and compare the resulting LR curves with those obtained from the observed database in order to detect any meaningful differences which might signal the inappropriateness of the allele independence assumptions. Such LR curves are very useful in illustrating the discriminating power of reported STR matches, as well as for placing reported LRs in context relative to the full LR distribution. Thus, we wish to check that there is no major discrepancy in practical terms between the observed LR curves and the set of curves resulting from datasets artificially constructed from the independence model, expression (1). See (10) for a more detailed discussion of the interpretation of Tippett diagrams.

Comparison of Members of the Same Subpopulation

Consider the case where offender and suspect are distinct members of the same racial group subpopulation. We take subpopulations to refer to groups of finite size whose members have a tendency to mate amongst themselves rather than completely randomly within the entire racial group. This pattern of breeding may cause allele distributions to vary across the different subpopulations. However, data relating to specific subpopulations are generally unavailable and naive use of the independence model via expression (1) with allele distributions estimated from the entire racial group will tend to overstate the strength of the DNA evidence when comparing individuals from the same subpopulation. In an attempt to remedy this situation and in the absence of appropriate subpopulation data, we may implement the formula of Balding and Nichols (11,12) which expresses the match probability in this case as a product across loci of single-locus match probabilities given by

$$\begin{cases} \dfrac{[2\theta + (1-\theta)\gamma_j][3\theta + (1-\theta)\gamma_j]}{(1+\theta)(1+2\theta)}, & \text{for the homozygote } (A_j, A_j) \\ \dfrac{2[\theta + (1-\theta)\gamma_j][\theta + (1-\theta)\gamma_k]}{(1+\theta)(1+2\theta)}, & \text{for the heterozygote } (A_j, A_k) \end{cases} \quad (4)$$

where θ denotes the coancestry between individuals in a subpopulation. We note that this approach is endorsed by the new NRC guidelines (4).

We thus have an explicit formula for the match probability associated with the profile $\mathbf{x}$ in terms of the two unknown quantities, $\theta = (\theta_1, \theta_2, \ldots, \theta_M)$, the set of coancestry coefficients at each of *M* loci, and γ, the set of racial group allele distributions. The full Bayesian approach then involves finding the posterior distribution for θ and γ given data, $\mathscr{D}$, from the suspect's racial group, $p(\theta, \gamma|\mathscr{D})$, with respect to which (4) is integrated to yield match probabilities and, by inversion, LRs. The method adopted here to implement this Bayesian approach in the absence of appropriate subpopulation data was originally proposed by Roeder et al. (13) and this was subsequently adapted and developed further in (3).

In order to specify the likelihood term, $p(\mathscr{D}|\theta, \gamma)$, we first introduce what we refer to as the *substructure model.* At each locus, a probability model (e.g., see (14)) is adopted which is often used to describe the process by which genotypes are generated within an inbreeding racial group subpopulation:

$$\begin{cases} \theta\gamma_j + (1-\theta)\gamma_j^2, & \text{for the homozygote } (A_j, A_j) \\ 2(1-\theta)\gamma_j\gamma_k, & \text{for the heterozygote } (A_j, A_k) \end{cases} \quad (5)$$

Note that we have made the further assumption that individuals within subpopulations mate at random, in which case the inbreeding coefficient in (5) equates to the coancestry, θ. Thus, the substructure model offers an alternative method for calculating the probability of profile $\mathbf{x}$ within a racial group subpopulation exhibiting allele distributions different from γ in $\mathscr{P}$; i.e., by multiplying together terms such as (5) across loci.

As in the previous section, we may then define the likelihood term, $p(\mathscr{D}|\theta, \gamma)$, to be a product of n profile probabilities, one for each individual in the database, using the substructure model via (5). In our earlier paper (3), we refer to this special case as the *profile-product likelihood* and θ may then be interpreted as a "combined" coancestry measure which applies to typical subpopulations of $\mathscr{P}$.

From Bayes' theorem in expression (2), given a prior distribution, $p(\theta, \gamma)$, the posterior distribution is then given by $p(\theta, \gamma|\mathscr{D}) \propto p(\mathscr{D}|\theta, \gamma)p(\theta, \gamma)$, which is a complicated expression in terms of θ and γ. Thus, the posterior match probability of any profile **x** cannot be found directly by integration of (4) for each locus. However, the *Gibbs sampler* (15) is a statistical simulation strategy which can be applied iteratively to generate a sample of values from virtually any distribution, such as a sample for θ and γ from the posterior distribution $p(\theta, \gamma|\mathscr{D})$. We may then estimate the value of any posterior quantity, $f(\theta, \gamma)$ (the match probability of **x**, say, using (4)), in the form of an average across this posterior sample. The Appendix of (3) provides a detailed description of the Gibbs sampling strategy which applies under adoption of the profile-product likelihood and this is implemented in the analyses of the following section.

Results

The original quadruplex data, typed at 1. VWA, 2. THO1, 3. F13A1, 4. FES, were compiled from a number of different sources as described in reference (2). Comparisons of allele distributions described in (1) show very little variation between samples within racial groups and, hence, the decision to combine the data into three databases representing the three racial groups. The datasets analyzed in this section are compiled from all complete 4-locus profiles contained in the source data; i.e., the final database sizes, n, are 1400 (Caucasian), 533 (Afro-Caribbean), 556 (Asian) and 2489 (mixed population). The combination of all three racial groups forms a single database representing a substructured mixed population and this will be useful for comparison purposes later in the section.

Investigation of the Independence Model Using LR Curves

We first check whether adopting the independence model to compute LRs is valid when comparing completely unrelated individuals. This is achieved, under the first criterion that investigates the effect on LR distributions, by implementing the LR curve comparisons described earlier.

Initially, we consider the plug-in approach where allele proportions in γ are fixed to be their empirical plug-in estimates in $\hat{\gamma}$. Figure 1 gives the LR curve plots under within- and between-person comparisons resulting from analysis of the Caucasian data. These are based on a sample of $T = 10\,000$ databases generated according to the independence model, which was found to be sufficient for adequate estimation of the distribution of resulting LR curves. The unbroken lines correspond to curves for the observed dataset $\mathscr{D}$ and the dotted lines represent upper 95th and lower 5th per-

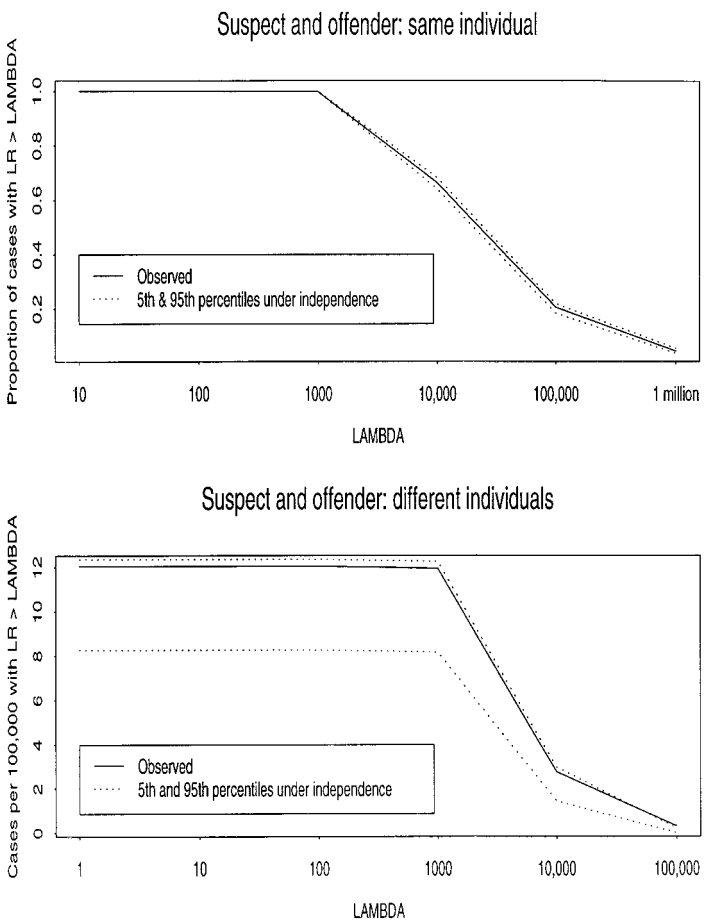

FIG. 1—*CAUCASIAN: comparison of observed LR curves (—) with the 5th and 95th percentile curves (...) constructed from a sample of* T *= 10 000 databases generated from the independence model.*

centile curves obtained from datasets generated according to the independence model; i.e., the dotted lines define an "envelope" within which LR curves corresponding to the artificially-constructed databases will lie 90% of the time.

The top graph presents LR curves in the case where suspect and offender are the same person. We see that the observed curve lies well within the dotted line envelope. The graph indicates that when comparing profiles from the same individual, we always expect to obtain a LR value greater than 1000 and, furthermore, in about 65% of cases, it is likely to be greater than 10 000. Arguably, the bottom graph corresponds to the situation of most interest to us; i.e., when suspect and offender are not the same person. In the observed database, 118 from a total of 979 300 between-person comparisons resulted in a match, which translates to just over 12/100 000 cases in which the LR > 0. We see that this conclusion is not atypical of databases constructed from the independence model. Furthermore, a LR value greater than 10 000 is only likely to be observed for about two or three cases in every 100 000 when the suspect is not the offender and the population satisfies the independence assumptions; this is in agreement with the interpretation resulting from the observed curve. It can be seen, therefore, that the discriminating power of STR matches under the quadruplex profiling system is clearly illustrated by investigation of the LR distributions and the conclusions drawn from studying the observed curve are practically indistinguishable from those exhibited by databases satisfying the independence model exactly.

Similar conclusions can be drawn from the graphs obtained for the Afro-Caribbean and Asian data, not shown here; e.g., observed LR curves lie within the 5th and 95th percentile dotted line bounds, although the LR distributions differ slightly in shape (see (2) for the observed LR curves). For reference, we note that additional analyses were carried out on "observed" datasets simulated from models in which substantial allele dependencies were induced and for which adopting the independence model would be inappropriate. The observed LR curves were located very clearly outside the independence envelopes in these cases.

Thus, our analyses suggest that adopting the simplifying assumptions of within- and between-locus independence of alleles, when comparing completely unrelated individuals, yields LR distributions which are robust to the levels of deviation from these assumptions that actually occur in practice; i.e., observed LR curves yield practically the same conclusions when used to establish the discriminating power of a match as LR curves typical of databases generated under the independence model. Furthermore, LR curve plots obtained using the Bayesian approach and expression (3) to calculate profile probabilities are virtually identical to those described above and, thus, fully support the robustness claims made under the plug-in approach.

Inference for the Coancestry Coefficient θ

In the first instance, we consider the case where allele proportions in γ are fixed to be their empirical estimates in $\hat{\gamma}$. We adopt a Beta(1.5,50) prior for θ_j at each locus j (3). This is a standard statistical distribution defined on the interval [0,1] which is unimodal and assigns zero density to a value of 0 or 1. It represents prior be-

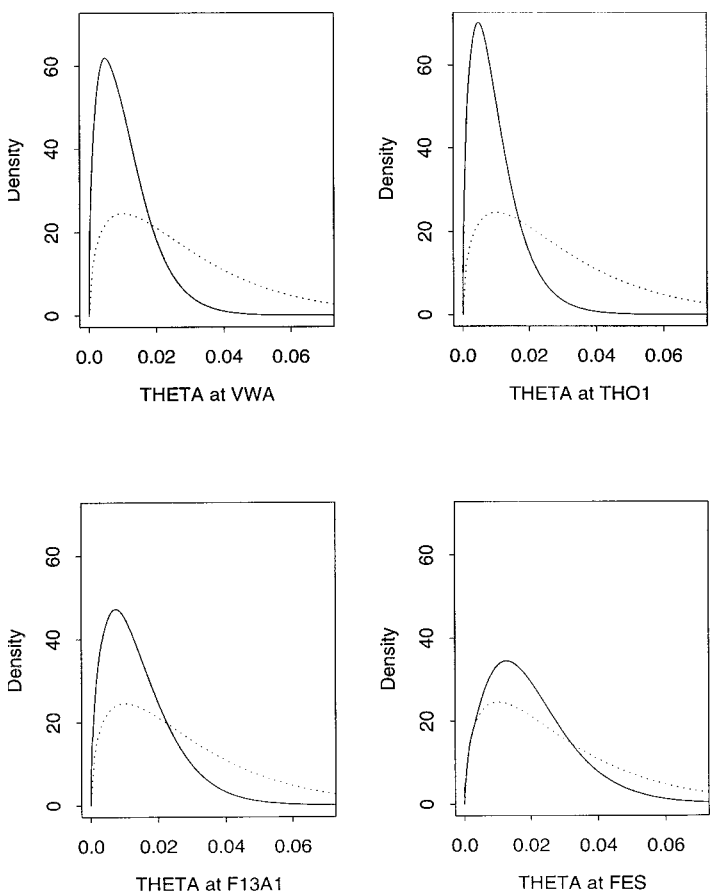

FIG. 2—*CAUCASIAN: comparison of prior (...) and posterior (—) distributions for θ.*

TABLE 1—*Prior and posterior distribution summaries for θ.*

Locus	5th Percentile	Median	Mean	75th Percentile	95th Percentile
			Prior		
All loci	0.0035	0.0233	0.0291	0.0401	0.0748
			Caucasian		
VWA	0.0017	0.0094	0.0107	0.0149	0.0240
THO1	0.0015	0.0083	0.0097	0.0133	0.0223
F13A1	0.0022	0.0126	0.0141	0.0196	0.0311
FES	0.0039	0.0186	0.0200	0.0273	0.0410
Average		0.0123	0.0136		
			Afro-Caribbean		
VWA	0.0029	0.0156	0.0175	0.0239	0.0389
THO1	0.0027	0.0159	0.0181	0.0252	0.0408
F13A1	0.0019	0.0112	0.0130	0.0179	0.0302
FES	0.0013	0.0081	0.0099	0.0138	0.0249
Average		0.0127	0.0146		
			Asian		
VWA	0.0035	0.0185	0.0205	0.0282	0.0448
THO1	0.0015	0.0095	0.0112	0.0156	0.0271
F13A1	0.0018	0.0118	0.0138	0.0190	0.0329
FES	0.0083	0.0313	0.0331	0.0440	0.0645
Average		0.0178	0.0197		
			Mixed		
VWA	0.0048	0.0158	0.0164	0.0213	0.0297
THO1	0.0060	0.0189	0.0195	0.0254	0.0351
F13A1	0.0166	0.0314	0.0316	0.0377	0.0473
FES	0.0081	0.0213	0.0217	0.0274	0.0369
Average		0.0219	0.0223		

liefs that the coancestry coefficient θ at any locus probably lies somewhere between 0 and 0.075, with density concentrated about median and mean values of 0.023 and 0.029, respectively. Prior specification was based partly on theory (i.e., the magnitude of coancestry values for close blood relatives) and partly on prior empirical evidence from other STR studies. Since higher θ values yield greater match probabilities in expression (4), it was deemed preferable (more *conservative*) to assume stronger prior beliefs about larger (and yet realistic) θ values than we truly believed before observing the data, since this will tend to yield weaker evidence against the suspect; i.e., we start from a position that is aimed at being defensible in court while not unrealistically detracting from the power of the STR profiling technique.

The Gibbs sampler was run for 30 000 iterations on each racial group database. Posterior distributions based on the Caucasian data and reconstructed from the entire posterior sample for θ are given in Fig. 2. These tend to be more concentrated about lower values for the θ_j's than the priors. This general observation is confirmed by inspection of the posterior distributions for all three racial groups as summarized in Table 1 and supports our claim that prior specifications were conservative. Furthermore, we may compare Table 1 with the analogous posterior distribution summaries presented in Table 3 of (13), noting that these authors perform analyses on the basis of VNTR (variable number tandem repeat) loci and adopt a tighter Beta(1,49) prior about the lower mean of 0.02 for each θ_j. Within the Caucasian group, for example, databases analyzed by Roeder et al. in (13) are 2000 to 3000 profiles in size, yielding posterior medians in the range 0.001 to 0.0035. These are much lower and more similar to values for coancestry measures observed elsewhere in the literature (1) than those obtained based on our Caucasian database of size 1400. Since allele mutation rates tend to be higher at VNTR loci than STR loci, lower θ values are to be expected, although this provides only a partial explanation for the differences. An additional factor is that the STR databases studied here do not have the power (in terms of volume of data) to completely overwhelm the conservative priors. Furthermore, by inspection of Table 1 in (1), which analyzes STR datasets similar to ours, we observe that lower levels of heterozygosity tend to correspond to racial group/locus combinations that exhibit higher θ_j values in Table 1 of this paper. This is to be expected since greater levels of population substructure yield greater levels of homozygosity.

This point is further supported by inspection of the posterior distribution summaries for the mixed population, where the database size in this case is 2489. We would expect the levels of subpopulation coancestry and inbreeding to be much greater within the mixed population than within each of its component racial groups. Thus, the posterior distribution of θ_j at each locus for the mixed population may be thought of as an "upper bound" for its constituent groups considered separately. Therefore, the high values yielded at the FES and VWA loci within the Asian group in particular might be considered a result of insufficient data.

To facilitate comparisons across loci and between populations, it is useful to have single posterior summaries of θ_j at each locus. We recommend the use of posterior means for reasons that will become clear in the next section. A single summary measure of the full posterior distribution for θ chosen to represent the general level of coancestry exhibited within a population is then given by the mean value of θ_j averaged across loci; this allows easy comparison with values obtained in other studies. From Table 1, the quadruplex data are suggesting values of approximately 0.014 (Caucasian), 0.015 (Afro-Caribbean) and 0.02 (Asian).

Calculation of LRs

Recall that we may express the match probability for two members of the same subpopulation in terms of θ and γ using the formula (4). As noted at the end of the previous section, the correct Bayesian approach evaluates the probability of a match **x** by aver-

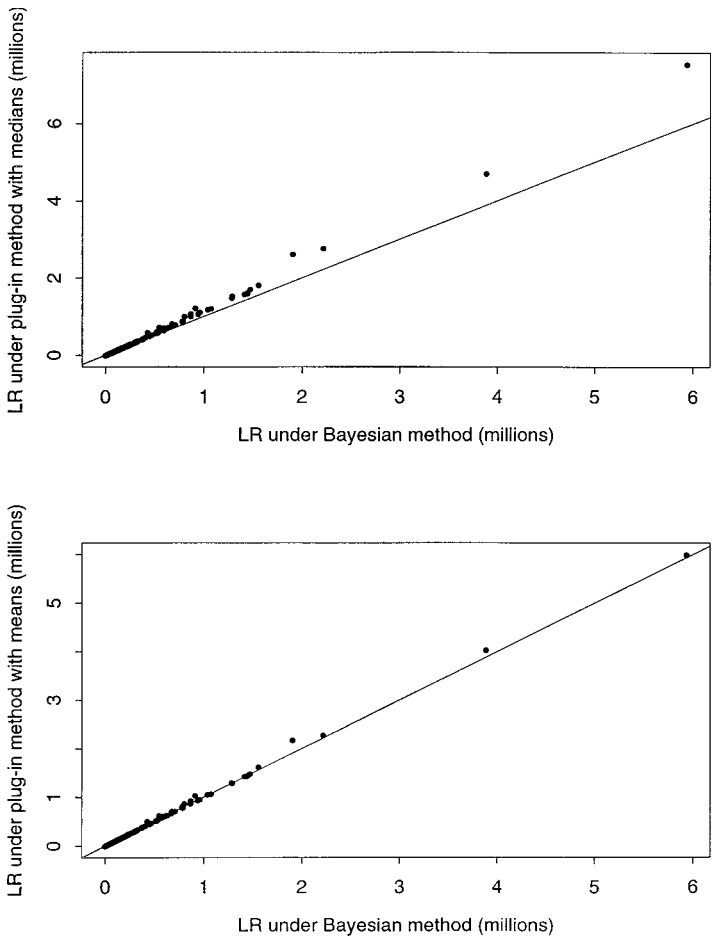

FIG. 3—*CAUCASIAN: scatterplots of LRs under the Bayesian vs. plug-in methods.*

aging (4) across the sample of θ values from the posterior distribution for each locus and then multiplying. Alternatively, a simple approximation substitutes a suitable posterior value for θ in the match probability formula—the so-called "plug-in" approach. To investigate the effect on individual LR values, we may compute the LR corresponding to each profile observed in a database under both methods and compare.

For the Caucasian data, Fig. 3 gives scatterplots of "the LR evaluated under the Bayesian approach" versus "the LR evaluated under the plug-in approach" for each profile in the database. It can be seen that LR values are practically indistinguishable from those obtained under Bayesian integration for the majority of profiles (i.e., points lie close to the "$x = y$" line) when posterior mean estimates, as opposed to posterior medians (13), say, are substituted for the θ_j's under the plug-in approach. Furthermore, we see that as matching profiles become rarer, absolute differences between LRs computed under the two methods will increase, although such discrepancies tend only to be large for very rare profiles composed largely of homozygotes. These results are confirmed by inspection of the most extreme cases where LRs are computed for the rarest (composed entirely of homozygotes) and commonest profiles identified under independence assumptions (see Table 2).

From the above analyses, the mean of θ at each locus contains all the information from the posterior distribution which is necessary for evaluating match probabilities and, thus, LRs, our ultimate aim. On this basis, the posterior mean would seem to represent a suitable summary measure for population coancestry levels.

TABLE 2—*LRs computed under the Bayesian and plug-in methods for the rarest and commonest profiles.*

Profile	Bayesian	Plug-in Estimates	
		Mean	Median
	Caucasian		
Rarest	2.659×10^{11}	8.354×10^{11}	2.032×10^{12}
Commonest	907	907	916
	Afro-Caribbean		
Rarest	8.218×10^{10}	3.620×10^{11}	1.090×10^{12}
Commonest	2114	2113	2155
	Asian		
Rarest	2.223×10^{10}	7.421×10^{10}	1.847×10^{11}
Commonest	1628	1626	1654

Results Under the Full Bayesian Approach

A full Bayesian analysis may be conducted where we represent our uncertainty about γ in addition to θ through a posterior distribution. With the adoption of flat (uniform) priors for the allele distribution at each locus in γ, which may be interpreted as representing prior ignorance about γ, the posterior distribution of θ is little changed from that given in Table 1 when $\gamma = \hat{\gamma}$. Furthermore, the posterior mean estimate of the allele distribution at each locus for each racial group is almost indistinguishable from its empirical estimate as given by $\hat{\gamma}$ and constructed from the observed databases. As in the calculation of LRs section, substitution of posterior mean

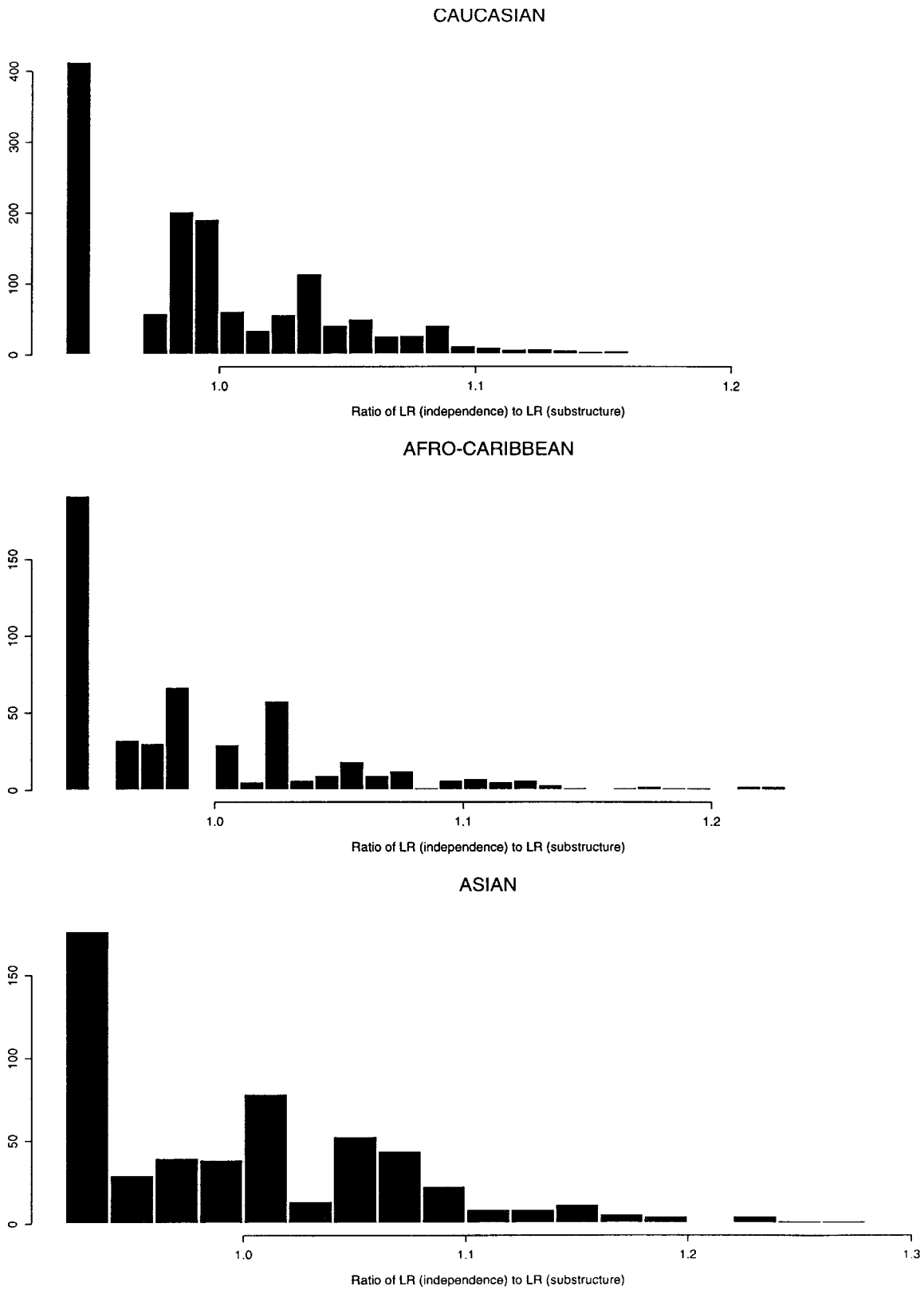

FIG. 4—*Histograms plotting the distribution of the ratio of LR values calculated under the independence vs. substructure model.*

estimates for both θ and γ in the formula (4) adequately approximates the full Bayesian match probability calculations.

Comparison of the Independence and Substructure Models

Previously, we introduced the substructure model (5) from which we constructed the likelihood term used to draw inference about θ values. It was noted then that this model could be adopted as a, possibly more realistic, alternative to the independence model (1) when evaluating STR profile probabilities; i.e., by incorporating the effects of population substructure through subpopulation coancestry levels. Thus, the substructure model (5) may be used in LR calculations when comparing individuals who are unrelated. The difference in LR values resulting from adoption of model (1) in place of the more realistic model (5) then provides a further criterion, in addition to the LR distribution comparisons, by which the appropriateness of allele independence assumptions in this case may be judged. Thus, for each profile observed in the racial group datasets, LRs may be evaluated using both (1) and (5), substituting $\hat{\gamma}$ for the allele distributions and posterior means estimated for the θ_j's as given in Table 1. For STR profiles contained in the quadruplex datasets, Fig. 4 shows that adopting the independence model yields LRs that tend to be less than 1.3 times their corresponding value under the substructure model. We note that within each dataset, there were 2 to 4 profiles which yielded a ratio greater than 1.3 but these were removed for the purposes of plotting the histograms. Furthermore, it can be seen that in more than 50% of cases, the independence model results in lower, and thus more conservative, LR values—these correspond to profiles that are largely heterozygous in nature. This is due to the fact that genotype probabilities are higher under model (1) as compared with model (5) for heterozygotes.

Thus, for substructure levels observed in our quadruplex data, the simple independence model provides an adequate approximation to the more realistic substructure model. This serves to support the evidence of the LR curve comparisons described earlier, i.e., the practical effect of observed levels of substructure on both individual LR values and population LR distributions computed under the independence model is negligible when comparing completely unrelated individuals.

Discussion

Bayesian analyses conducted on the quadruplex datasets discussed in this paper serve to validate their use in quantifying STR evidence for identification purposes. The approach adopted to evaluate LR values when a match occurs necessarily changes according to the assumed relationship between suspect and offender if they are not the same person.

In the first case, we might consider suspect and offender to be completely unrelated individuals whose STR profiles happen to match. We are then reduced to evaluating a proportion for the matching profile within the racial group of possible offenders. This can be done via the independence model, even though the idealized conditions necessary for exact allele independence never hold in real populations. It has been shown that distributions of LRs estimated using the independence model for profiles in the three observed racial groups are practically the same as those typically seen in equivalent populations within which the independence model is truly valid. Furthermore, individual LRs are robust to the independence assumptions when compared with what might be considered a more realistic substructure model. All the analyses provide support for use of the simplified model in casework.

When circumstances of a case make it reasonable to assume the suspect and offender are members of the same racial group subpopulation, the NRC report (4) recommends the adoption of Balding and Nichols' (11,12) conditional match probability expression (4). This necessitates making inference about typical subpopulation coancestry measures, θ. Adopting a Bayesian approach is useful in allowing the incorporation of prior information about plausible θ values based on other studies of STR data as well as population genetics theory. General levels of subpopulation coancestry estimated from the quadruplex data were 0.014 (Caucasian), 0.015 (Afro-Caribbean) and 0.02 (Asian). Furthermore, simply substituting posterior mean values of θ in the match probability formula (4) was shown to provide a simple and adequate approximation to full Bayesian integration for use when LRs are calculated in practice. Therefore, our analyses based on identifying general coancestry levels within the main UK racial groups suggest that routine use of a value of $\theta = 0.03$ in LR calculations, as recommended in the NRC guidelines (4), will tend to understate the DNA evidence in cases where the suspect and offender are assumed to be distinct members of a typical UK subpopulation.

References

1. Gill P, Evett IW. Population genetics of short tandem repeat (STR) loci. Genetica 1995;96:69–87.
2. Evett IW, Gill PD, Scranage JK, Weir BS. Establishing the robustness of STR statistics for forensic applications. Am J Hum Genet 1996;58:398–407.
3. Foreman LA, Smith AFM, Evett IW. Bayesian analysis of DNA profiling data in forensic identification applications (with discussion). J Roy Stat Soc A 1997;160:429–69.
4. National Research Council. The evaluation of forensic DNA evidence. Washington DC: National Academy Press, 1996.
5. Roeder K. DNA fingerprinting: a review of the controversy (with discussion). Stat Sci 1994;9:222–78.
6. Kimpton C, Fisher D, Watson S, Adams M, Urquhart A, Lygo JE, et al. Evaluation of an automated DNA profiling system employing multiplex amplification of four tetrameric STR loci. Int J Leg Med 1994;106:302–11.
7. Evett IW. Evaluating DNA profiles in a case where the defence is `It was my brother'. J Forensic Sci 1992;32:5–14.
8. Zaykin D, Zhivotovsky L, Weir BS. Exact tests for association between alleles at arbitrary numbers of loci. Genetica 1995;96:169–78.
9. Evett IW, Scranage J, Pinchin R. An illustration of the advantages of efficient statistical methods for RFLP analysis in forensic science. Am J Hum Genet 1993;52:498–505.
10. Evett IW, Buckleton JS. Statistical analysis of STR data. In: Carracedo A, Brinkmann B, Bar W, editors. Advances in forensic haemogenetics 6. Springer, 1996:79–86.
11. Balding DJ, Nichols RA. DNA profile match probability calculations: how to allow for population stratification, relatedness, database selection and single bands. Forensic Sci Int 1994;64:125–40.
12. Balding DJ, Nichols RA. A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity. Genetica 1995;96:3–12.
13. Roeder K, Escobar M, Kadane JB, Balazs I. Measuring heterogeneity in forensic databases using hierarchical Bayes models. Pittsburgh (PA): Carnegie Mellon University; 1994 Technical Report.
14. Weir BS. Genetic data analysis. Sunderland, MA: Sinauer Associates, 1990.
15. Smith AFM, Roberts GO. Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. J Roy Stat Soc B 1993;55:3–23.

Additional information and reprint requests:
Dr. Lindsey A. Foreman
The Forensic Science Service
Metropolitan Laboratory
109 Lambeth Road
London
SE1 7LP
UK

ERRATA/CORRECTIONS

We have identified a number of instances in which the authors of work published in the Journal of Forensic Sciences have miscited papers originally published in the Journal of the Forensic Science Society as having been published in the Journal of Forensic Sciences.

The known instances of this error for volume 44 of the Journal of Forensic Sciences are detailed/corrected below. We have not checked other volumes for similar errors. The Journal of Forensic Sciences regrets these errors.

Since 1995 (Volume 35), the Journal of the Forensic Science Society has been published under the title "Science and Justice."

The editors of both journals take this opportunity to remind authors of the necessity for ensuring the accuracy of the references they cite in manuscripts submitted for publication. The Instructions for Authors of both journals make it clear that accuracy of reference citation is the responsibility of authors, and good scholarship demands attention to this matter.

A. R. W. Forrest	R. E. Gaensslen
Editor, Science and Justice	Editor, Journal of Forensic Sciences

The journal citation in reference 7 in Foreman LA, Smith AFM, Evett IW. Bayesian validation of a quadriplex STR profiling system for identification purposes. should read: J Forensic Sci Soc 1992;32:5–14.

The journal citation in reference 5 in Bourel B, Hedouin V, Martin-Bouyer L, Becart A, Tournel G, Deveaux M, Gosset D. Effects of morphine in decomposing bodies on the development of Lucila sericata (Diptera: Calliphoridae). should read: J Forensic Sci Soc 1991;31:469–72.

The journal citation in reference 8 in Hedouin V, Bourel B, Martin-Bouyer L, Becart A, Tournel G, Deveaux M, Gosset D. Determination of drug levels in larvae of Lucila sericata (Diptera: Calliphoridae) reared on rabbit carcasses containing morphine. should read: J Forensic Sci Soc 1994;34:95–7.

The journal citation in reference 15 in Hedouin V, Bourel B, Martin-Bouyer L, Becart A, Tournel G, Deveaux M, Gosset D. Morphine perfused rabbits: A tool for experiments in forensic entomotoxicology. should read: J Forensic Sci Soc 1991;31: 469–72.

The journal citation in reference 10 in McDermott SD, Willis SM, McCullough JP. The evidential value of paint. Part II. A Bayesian approach. should read: J Forensic Sci Soc 1992;32: 333–48.

The journal citations in references 4 and 5 in Infante F, Dominguez E, Trujillo D, Luna A. Metal contamination in illicit samples of heroin. should read for 4: J Forensic Sci Soc 1979;19:203–9. and for 5: J Forensic Sci Soc 1980;20:177–81. [in reference 5 only the volume number is miscited]. And in both references, the lead author's name is "Joyce JR."

The journal citation in reference 1 in Savolainen P, Lundeberg J. Forensic evidence based on mtDNA from dog and wolf hairs. should read: J Forensic Sci Soc 1988;28:335–9.

The journal citation in reference 1 in Kupfer DM, Chaturvedi AK, Canfield DV, Roe BA. PCR-based identification of postmortem microbial contaminants—A preliminary study. should read: J Forensic Sci Soc 1968;8:73–6.

In every instance cited above, future citations of the J Forensic Sci papers containing the errors should contain the following: [published erratum appears in J Forensic Sci 2001 Jan;46(1)] immediately following the article title and before the journal citation, in accordance with the Uniform Requirements for the Submission of Manuscripts to Biomedical Journals style.