Available online at www.sciencedirect.com
ScienceDirect
Forensic Science International xxx (2007) xxx–xxx
Forensic Science International
www.elsevier.com/locate/forsciint

DNA-testing for immigration cases: The risk of erroneous conclusions

Andreas O. Karlsson [a,*], Gunilla Holmlund [a], Thore Egeland [b], Petter Mostad [c]

[a] *The National Board of Forensic Medicine, Department of Forensic Genetics and Forensic Toxicology, Artillerigatan 12, SE-581 33 Linköping, Sweden*
[b] *Department of Medical Genetics, Ullevaal University Hospital, NO-0407 Oslo, Norway*
[c] *Department of Biostatistics, Institute of Basic Medical Sciences, University of Oslo, P.O. Box 1122, Blindern, NO-0317 Oslo, Norway*

Received 30 August 2006; received in revised form 27 November 2006; accepted 3 December 2006

Abstract

Making the correct decision based on results from DNA analyses and other information in family reunification cases can be complicated for a number of reasons. These include stratified populations, cultural differences in family constellations, families with different population origin, and complicated family relations giving complex pedigrees. The aim of this study was to analyze the risk of erroneous conclusions in immigration cases and to propose alternative procedures to current methods to reduce the risk of making such errors. A simulation model was used to study different issues. For simplicity, we focus on cases which can be formulated as questions about paternity. We present an overview of error rates (of falsely included men as the true father and of falsely excluded true fathers) for fairly standard computations, and we show how these are affected by different factors. For example, adding more DNA markers to a case will decrease the error rates, as will the inclusion of more children. We found that using inappropriate population frequency databases had just minor effects on the error rates, but the likelihood ratios varied from an underestimation of 100 times up to an overestimation of 100,000 times. To reduce the risk of falsely including a man related to the true father we propose a more refined prior including five hypotheses instead of the two normally used. Simulations showed that this method gave reduced error rates compared with standard computations, even when the prior does not exactly correspond to reality.

Keywords: Immigration casework; Bayesian; Mutation model; Simulation

1. Introduction

The use of DNA analysis in forensic casework has revolutionized the area of forensic science. Suspects can be linked to crime scenes, victims of mass disasters such as airplane crashes can be identified and questions of disputed paternity can be solved [1,2]. DNA can also be used as a tool in immigrations cases, which is the focus of this paper.

Immigration casework involves family reunification and often consists of a man who would like to be reunified with his wife and children. In such cases, the question is whether a given man, the alleged father (AF) is the true father (TF) of a number of children. The maternity of the alleged mother is usually not questioned. In general, all stated relationships could be questioned, and a range of different pedigrees could be used as hypotheses. Methods presented in this paper could be directly extended to such situations. However, as most cases in practice seem adequately represented by the paternity perspective,[1] this will be the focus of our paper. Treating immigration cases as paternity cases they have, however, some characteristics of their own, like frequently involving populations where knowledge of allele frequencies may be inadequate. Also, there may be varying degrees of consanguinity making it reasonable to analyze alternative pedigrees.

Our aim was to study the risk of erroneous conclusions in immigration cases and to propose alternatives to current methods to reduce the risk of making such errors.[2] In a paternity case a decision can comprise two types of errors. Here we define them as ‘‘exclusion errors’’ and ‘‘inclusion errors’’, meaning a false exclusion, respectively, a false inclusion, of the

[1] Note that cases where maternity but not paternity is in question are entirely similar, by symmetry.

[2] Note our use of the term error: this paper does not discuss DNA-typing errors, calculation or typing errors or the like. An error is for us an erroneous conclusion in the question underlying the immigration case.

* Corresponding author. Tel.: +46 13 31 58 70; fax: +46 13 13 60 05.
E-mail address: andreas.karlsson@rmv.se (A.O. Karlsson).

0379-0738/$ – see front matter © 2007 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.forsciint.2006.12.015

AF from paternity. There will always be a balance between these errors and the goal is to reduce them simultaneously. These error rates together with likelihood ratios are used for measuring the impact of different parameters in the statistical computations.

An important issue in all paternity cases is how to handle inconsistencies between the profiles of an AF and a child. This issue is not specific neither for paternity nor immigration casework, but since we in this paper deal with describing and evaluating the possibilities of misclassification in immigration cases, mutations might have an impact. One approach to deal with this is to use a probabilistic mutation model. Different models have been proposed [3–5] but none have yet been generally adopted. An alternative way to account for inconsistencies is to exclude them from the likelihood calculations and instead set a limit of a maximum of one or two inconsistencies between the profiles of an AF and a child (later referred as 1-incon and 2-incon, respectively) before rejecting paternity. The rationale behind the approach would be that it should have comparable error rates with the first alternative, but without the need to decide on a mutation model. As different laboratories use different approaches in cases including inconsistencies, it is relevant to study our main questions in the context of these different approaches.

In addition to studying levels of error rates in standard cases, we also study how these rates are affected by a number of factors, such as the number of DNA markers used, the number of children involved, and the use of inappropriate population databases. Finally, we propose and study a computational method that explicitly takes into account the possibility of the AF being a close relative of the TF.

All studies were done using simulations. Simulation of families and their DNA profiles gives the opportunity to rather simply investigate different issues and also test the impact of changing the model and thus the influence of different parameters. This can ensure robustness of our results in relation to uncertainties about true population frequencies, possible familial relations, etc. With the limited number of thoroughly investigated real cases, it is difficult to see an alternative to using simulations.

2. Materials and methods

Different methods for paternity calculations can most easily be understood in a Bayesian framework [6,7]. Let H_0 be the hypothesis that AF is the TF, and let H_1 be the complementary possibility that he is not. Assuming that the likelihoods $LR = P(\text{data}|H_0)/P(\text{data}|H_1)$ [3] and $P(\text{data}|H_1)$ can be calculated, and writing $LR = P(\text{data}|H_0)/P(\text{data}|H_1)$ for the likelihood ratio, Bayes formula on odds form gives

$$\frac{p_1}{1-p_1} = \text{LR}\frac{p_0}{1-p_0}$$

or

$$p_1 = \frac{\text{LR}(p_0/(1-p_0))}{\text{LR}(p_0/(1-p_0))+1}$$

where p_0 is the prior probability for H_0 and p_1 is the corresponding posterior probability $P(\text{data}|H_1)$. When we make the assumption $p_0 = 0.5$, we define $W = p_1$, and we get the Essen–Möller's formula

$$W = \frac{\text{LR}}{\text{LR}+1} = \frac{\text{PI}}{\text{PI}+1}$$

where PI is the paternity index. We see that W, PI and LR are all directly related, and for any fixed p_0, there is a direct relationship between LR and P_1. Below, we will fix $p_0 = 0.5$.

In practice, results from paternity calculations are most often used to make a decision: AF is declared as the TF, or not. It is then important to avoid both that the AF is falsely excluded as TF (exclusion error), and that the AF is falsely included as TF (inclusion error). Let C_I and C_{II} be the ''costs'' of these two types of errors. The expected cost of rejecting AF as TF is then equal to C_I times the probability that H_0 is true, and the posterior expected cost becomes $p_1 C_I$. Similarly, the posterior expected cost of accepting AF as TF is $(1-p_1)C_{II}$. It is clear that the expected posterior cost is minimized when we declare AF as TF whenever

$$p_1 C_I \geq (1-p_1)C_{II}$$

or equivalently

$$p_1 \geq \frac{1}{(C_I/C_{II})+1}$$

Below, we will use the assumption that $C_I = C_{II}$, corresponding to a cutoff value for p_I at 50%. Using this cutoff will then minimize the expected sum of the error rates.

It remains to be discussed how the likelihoods $P(\text{data}|H_0)$ and $P(\text{data}|H_1)$ can be computed. In practice, we must make simplifying assumptions. In this paper, we will use the following as our basic computational method, and an example of a fairly standard procedure: When computing $P(\text{data}|H_0)$, we assume that TF is a completely unrelated person, and that, when more than one child is involved, he is the father of all the children. A Swedish allele database ($n = 300$ individuals) is used to estimate frequencies. New, not earlier seen alleles are added to the frequency databases with a default frequency of $5/2n$, where n is the number of individuals in the database [8]. Calculations of likelihoods are done using the Familias program (http://www.nr.no/familias, [9]) with mutation models explained later. We will also consider the common procedure where loci with an inconsistency between the AF and the child are simply removed from the computations. If the number of inconsistent loci is above a maximum of either one or two loci, the AF is declared not to be the TF; otherwise, the results from the reduced computations are used. This avoids the specification of a mutation model.

If the likelihood computations correspond directly to (simulated) reality, we saw above how using a cutoff for p_I at 50% would minimize the sum of the error rates. However, in what we will use as our two hypotheses standard procedure, we use the cutoff 99.99%. The reason is that one suspects there are cases where a close relative of TF is declared to be the true father. In such cases, the LR can often be high, although usually lower than when AF is the TF. The high cutoff value is meant to compensate for this.

However, when there is prior knowledge that there is a possibility for a close relative of the AF to be the TF, and if it is possible to formulate this knowledge into a precise prior, it is better to base the decision on the actual posterior for such a model, and use a cutoff at 50%, than to use an ad-hoc adjustment. Thus, we propose the following method:

We split the hypothesis H_1 into four different hypotheses:

H_{1a}. TF is the brother of AF.

H_{1b}. TF is the father of AF.[4]

H_{1c}. TF is the half-brother of AF.

[3] The notation $P(\text{data}|H_0)$ means: the probability of observing the given data given that the hypothesis H_0 is true.

[4] The alternative that TF is the son of AF is not included separately, as it would give exactly the same likelihood as when TF is the father of AF, except for some very small differences with some mutation models.

Please cite this article in press as: A.O. Karlsson et al., DNA-testing for immigration cases: The risk of erroneous conclusions, Forensic Sci. Int. (2007), doi:10.1016/j.forsciint.2006.12.015

H_{1d}. TF is unrelated to AF.

The likelihood $P(\text{data}|H_1)$ is then computed as

$$P(\text{data}|H_1) = 0.25P(\text{data}|H_{1a}) + 0.25P(\text{data}|H_{1b}) + 0.25P(\text{data}|H_{1c}) + 0.25P(\text{data}|H_{1d})$$

The weights 0.25 above are arbitrarily chosen, and could be more accurately set in actual cases. Our goal here is to show that even using such arbitrarily chosen weights results in a better performance than the ad-hoc adjustment of the cutoff value used in the two hypotheses standard procedure.

2.1. Simulation study

To investigate the sizes of error rates, and to study how these rates are affected by discrepancies between the simulated reality and assumptions used in the likelihood computations, we performed a simulation study. STR data from 15 loci, including the 13 CODIS core loci plus D19S433 and D2S1338 (included in the Identifiler kit, Applied Biosystems) were used for the simulation of DNA profiles. In some examples, profiles were generated with either 20 or 25 loci. This was done by reusing frequencies from 5 or 10 of the loci mentioned above. Generally, a Swedish allele database ($n = 300$ individuals) was used, but we also used an Iranian database ($n = 150$, [10]), a Somalian database ($n = 97$, [11]) and a Rwandan database ($n = 124$, [12]) to study the impact of using inappropriate allele databases in the computations. In the simulation of a founder DNA profile the alleles at each locus were randomly chosen based on the observed allele proportions in the database. An offspring profile (a child) was generated using Mendelian heritage based on the profile of the mother and the father. Mutations were allowed to occur when an offspring profile was simulated. For this a decreasing mutation model was used [13] with a μ-gen[5] of 0.1% per meiosis, locus, and generation (based on calculations from [14]). In a test of the robustness of computations using mutation models, we also used μ-gen of 0.05 and 1%. Ninety percent of the mutations consisted of ±1 repeat length differences and the rest were ±2 repeats length mutations [5].

In all simulations the profiles of a mother, a father, and one or two children were generated, and additionally a brother of the TF, a father of TF, a half-brother of TF and a, to the TF, unrelated man. In the standard examples the TF and the mother were assumed to be unrelated. When the impact of a relation between the father and the mother was tested, they were simulated as first cousins.

When simulating data under the H_1 hypothesis, the four hypotheses H_{1a}, H_{1b}, H_{1c}, and H_{1d} were used. In general, they were all assumed to be equally probable, so that the error rate for inclusion errors could be computed as the unweighted average rate of inclusion of AF as the TF under each of the hypotheses H_{1a}, H_{1b}, H_{1c}, and H_{1d}. However, robustness of the results was also considered by using weighted averages of these rates.

For each tested issue, 10,000 profiles were simulated for each person (TF, brother of TF, father of TF, half-brother of TF and an unrelated man).

3. Results

3.1. Standard examples

Exclusion and inclusion error rates for standard trios (a mother, child and AF), computed with Swedish allele data, are shown in Fig. 1. The exclusion error rate increased with an increasing cutoff level and the error rates of relatives or unrelated being computed as the TF decreased as the cutoff value increased. The most frequent inclusion error was the case where the brother of the TF or the father of the TF were included as the true father. The half-brother and the unrelated man were very seldom included as the TF. The sum of the exclusion and the inclusion error was the smallest between the cutoff values 99.9 and 99.99% (Fig. 1).

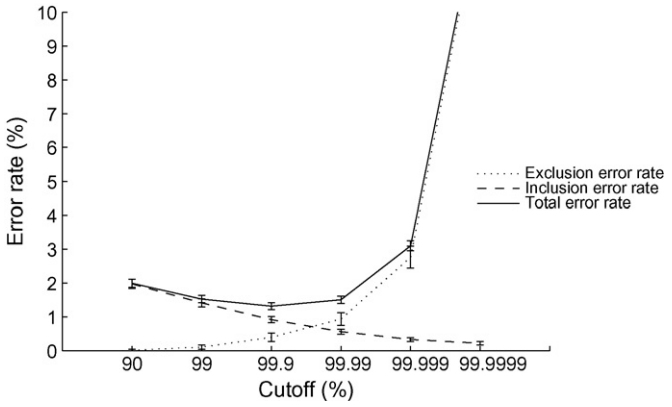

Fig. 1. Error rate as a function of the cutoff value. The exclusion error, the inclusion error and the total error were calculated for standard trios (mother, child and AF). The bars represent 95% confidence intervals for simulation uncertainty.

Adding more genetic markers to these standard cases did lower the error rates (Fig. 2), especially the exclusion error which decreased more than the inclusion error (data not shown). If one more child was added to the standard trio the total error rate decreased almost 10 times, from 1.5 to 0.18% with 99.99% cutoff level (Fig. 2).

The standard trios did not include any kinship relation between the mother and the AF. In a separate test where the father and mother were simulated to be first cousins there was no, or very little, influence on the error rates (Fig. 2).

3.2. Impact of using or not using a mutation model

The computed probabilities were, in the standard examples, based on a decreasing mutation model with the same μ-gen and μ-comp (0.1%). The obtained results could in theory be due to the use of an incorrect mutation model and also in real cases it is not possible to select the true mutation rate for the likelihood calculation. Only minor changes of the error rates occurred with altered μ-gen (0.05 and 0.1%) and μ-comp (0.05, 0.1, 0.5 and

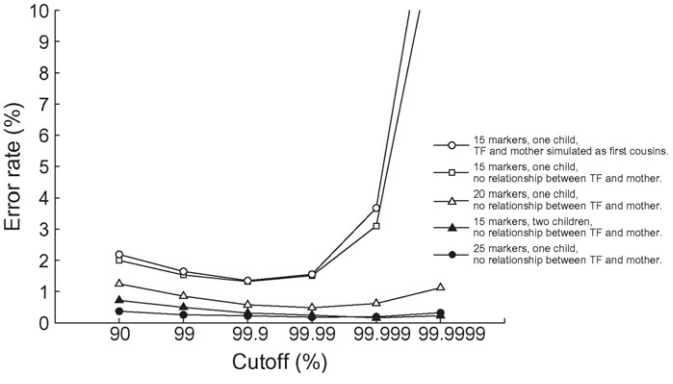

Fig. 2. The total error rates for trios with different conditions.

[5] The term μ-gen denotes the mutation rate used in the simulation of profiles, whereas the later used term μ-comp denotes the mutation rate in likelihood computation at a locus with an inconsistency between the AF and the child.

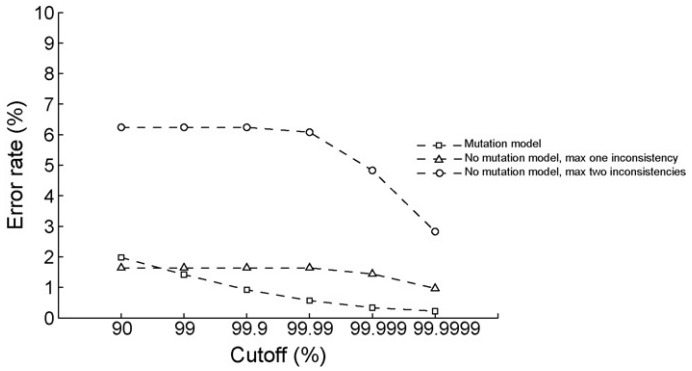

Fig. 3. The inclusion error rates from the different cases where a mutation model was or was not used.

1%) (data not shown). When μ-gen was set to 1% for all markers the total error rates increased with more than 50% (99.99% cutoff) using the same μ-comp as above.

We also used a standard trio to test how the error rates behaved when loci with inconsistencies were excluded from the likelihood computations. The simulations showed that the inclusion error rates became much higher using no mutation model. Using 99.99% as a cutoff the inclusion error rates were 6, 1.6 and 0.6% for the cases with 1-incon, 2-incon and a mutation model, respectively (Fig. 3). The exclusion error showed an opposite pattern with low error rate when loci with an inconsistency were excluded.

3.3. The effect of inappropriate allele databases

DNA profiles from families with a non-Swedish origin were simulated to illustrate the impact of using an inappropriate allele database for the likelihood calculations. Using Swedish allele frequencies, instead of the correct one, generally overestimated the likelihood ratios (medians: 4, 20 and 200 times; Iranian, Somalian and Rwandan allele frequencies, respectively) (Fig. 4). It is, however, important to notice that the variance was quite large. For example, in the Rwanda case the central 95% of the simulations covered the interval from an underestimation of 100 times up to an overestimation of almost 100,000 times of the ''correct'' LR.

The use of an inappropriate allele database did not, however, have that much influence on the error rates. More relatives of the TF were included as the TF using Swedish allele data compared with the correct allele frequencies. For instance, 2.5 times more brothers were included as the TF using Swedish (3.1% included brothers), instead of Rwandan (1.2%) allele frequencies using 99.99% as a cutoff. The exclusion error was lowered using an inappropriate frequency database making little change in the total error rate.

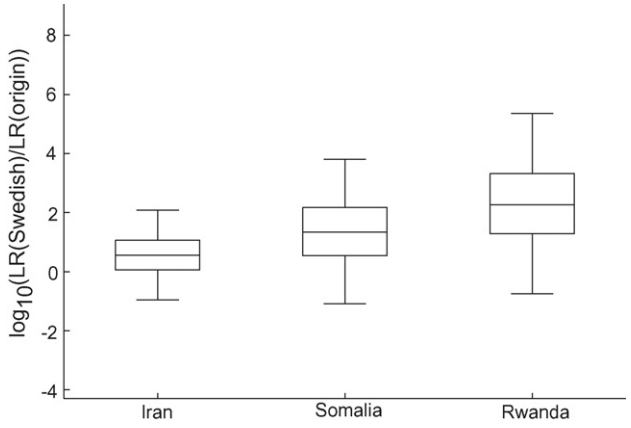

Fig. 4. Box plot showing the ratio of the likelihood ratios obtained with Swedish allele frequencies and the appropriate allele frequencies using simulated profiles with either Iranian (median: ~4 times), Somalian (~20 times) or Rwandan (~200 times) origin. The likelihood ratio was computed with the alleged father as the true father. The boxes have lines at the lower quartile, median, and upper quartile values. The whiskers are lines extending from each end of the boxes to show the extent of the rest of the data (covering 95%). The whiskers value used was one.

3.4. Five hypotheses model

As explained in the materials and methods sections, we can use a more refined prior, using five different pedigree hypotheses. This approach reduced both the exclusion and the inclusion error rates both when the hypotheses were considered individually and all together (Table 1, Fig. 5). Using appropriate cutoff values (99.99% for the two hypotheses model and 50% for the five hypotheses model) the total error rate was significantly reduced.

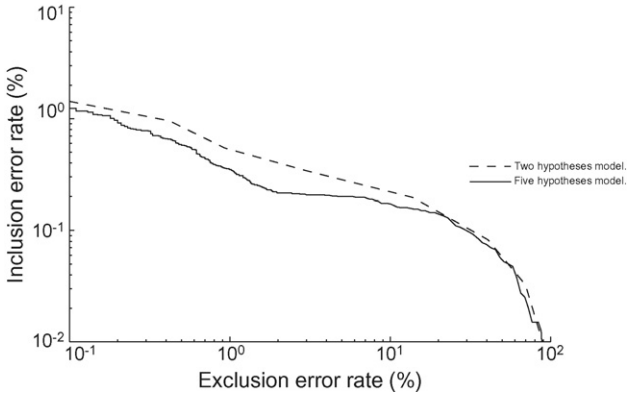

Fig. 5. Receiver operating characteristics (ROC) of the relation between the exclusion and the inclusion error rates using the two hypotheses standard model and the five hypotheses model.

Table 1
Error rates using the standard two hypotheses model and the five hypotheses model

Simulated alleged father	Error rate (%, (S.D. (95%)))		*P*-Value
	Two hypotheses model (99.99% cutoff)	Five hypotheses model (50% cutoff)	
True father (TF)[a]	0.94 (0.19)	0.65 (0.16)	0.02
Brother of TF[a]	1.22 (0.22)	0.92 (0.19)	0.04
Father of TF[a]	1.05 (0.20)	0.96 (0.19)	0.50
Half-brother of TF[a]	0.01 (0.02)	0.03 (0.03)	–
Unrelated to TF[a]	0	0	–
Exclusion error[a]	0.94 (0.19)	0.65 (0.16)	0.02
Inclusion error (mean)[b]	0.57 (0.07)	0.48 (0.07)	0.07
Total error[c]	1.51 (0.11)	1.11 (0.09)	0.006

[a] Based on 10,000 simulations.
[b] Based on 40,000 simulations.
[c] Based on 50,000 simulations.

4. Discussion

We used pedigrees and simulated DNA profiles to investigate inclusion/exclusion error rates, the impact of different mutation models and inappropriate databases in immigration casework statistics. We also suggest the use of a prior with a five hypotheses model for probability computations. The use of DNA analysis in family reunification is increasing and an evaluation of the statistical methods used is important. Our results apply to the overall operating characteristics, i.e., how certain rules and procedures affect error rates on an average.

The two types of error rates investigated consisted of TF excluded as the TF (exclusion error) and a non-true father included as the TF (inclusion error). The weighing between exclusion and inclusion errors is ultimately a legal and political question. Traditional methods tend to result in roughly minimal sum of error rates for the two types of errors under some assumptions, indicating that in practice the two types of errors have been considered equally costly. However, the ''cost'' of splitting up a family because of genetical coincidences could easily be considered much higher than the cost to society of admitting some extra persons not legally entitled to entry.

For the inclusion error we chose to report the mean of the different errors these rates consist of. We then assume that all the different hypotheses H_{1a}–H_{1d} are equally likely to occur when the AF is not the TF, an assumption most probably not true, but since we do not know their actual proportions we chose to treat them equally.

Simulating profiles with 25 STR markers, instead of the 15 markers, greatly reduced the error rates, especially the exclusion error. Typing of 25 different loci is today possible, but a reliable mutation model is required, since the chance of mutations increases as more markers are used. Furthermore, more markers will also increase the possibility of being linked or to be in linkage disequilibrium, which will make it hard to use the product role for independent markers.

Families with more than one child included in the pedigrees have lower error rates, but in these cases there are some important special considerations. For instance, the number of possible pedigrees will expand rapidly with the number of children, as the children may have different fathers.

To account for inconsistencies the use of a mutation model was in most cases better than not to use one even if μ-gen and μ-comp differed. As mentioned above, mutations are not of special interests in immigration casework but since these have an impact on inclusion/exclusion error rates the data from such simulations was presented. Our aim of this study was, however, not to propose a mutation model, but to see how the error rates were affected by different mutation rates excluding the risk that our results are only valid with the mutation model we have chosen to use.

In immigration casework we often have to rely on inappropriate allele frequency databases, an issue that has been discussed before [15] and also recently during the identification of the tsunami victims [16]. However, our calculations using different databases showed limited effects on the error rates although the impact on computed likelihoods could be quite large in individual cases. This difference in likelihood ratios is most probably due to the frequency of the paternal allele being typically lower in the Swedish database compared with a more appropriate one. Pronounced differences will also occur if an allele has never been observed in the database. In a simple case where the paternal allele is known the ratio (using Swedish allele frequencies compared with the ''real'') between the likelihood ratios could be given as:

$$\frac{\mathrm{LR_{Swe}}}{\mathrm{LR_{Correct}}} = \frac{p_i^{\mathrm{correct}}}{p_i^{\mathrm{swe}}}$$

where p_i^{correct} is the paternal allele frequency, or average if ambiguous, for locus i in the correct population and p_i^{swe} is the frequency, or average if ambiguous, in the Swedish database. The total ratio will then be

$$\frac{\mathrm{LR_{Swe}}}{\mathrm{LR_{Correct}}} = \prod_{i=1}^{15} \frac{p_i^{\mathrm{correct}}}{p_i^{\mathrm{swe}}}$$

For example if a number of alleles are relatively common in a given population and rare, or missing, in the Swedish population the ratio between the likelihood ratios will be highly affected. The opposite is less likely to occur and thus the likelihood ratio will be greater than one.

Please cite this article in press as: A.O. Karlsson et al., DNA-testing for immigration cases: The risk of erroneous conclusions, Forensic Sci. Int. (2007), doi:10.1016/j.forsciint.2006.12.015

Changing (increasing) the cutoff value can be a good idea in such cases, as it restores a more even balance between exclusion and inclusion error rates, and lowers the total error rate. However, the total error rate will still be higher than when using correct allele frequencies. The use of a weighted mean of calculated likelihoods, based on frequencies from different population databases, has been proposed for uncertain population affiliation [17].

The Swedish Migration Board claims that there are cases where the AF has been shown, by other means, to be the brother or some other close relative to the TF. Simulations confirmed that such a sibship has a considerable impact, especially in the context of using an inappropriate allele database and no mutation model. Adding more hypotheses, consisting of relatives to the TF, slightly reduced both the exclusion and the inclusion errors. Since the posteriors for the three cases of close relatives of TF (brother, father and half-brother of TF) are similar, we considered reporting the sum of the four ''non-paternity'' hypotheses (H_{1a}–H_{1d}) probabilities as a standard. This sum is based on the fact that if the AF is not the TF one of the other four hypotheses is true with equal probabilities. The correct proportions of these alternative hypotheses are not known and the used approach is fair for illustration.

The laboratory work does not differ between immigration casework and paternity testing. There are however differences in how to transform the DNA information into probability numbers. Immigration casework is about collecting information such as knowledge about the family constellation, population affiliation and also including as many more children as possible. In many cases these opportunities are not available. Our study showed, however, that although information is missing, these cases can often be correctly interpreted if valid methods and strategies are used for the likelihood computations.

Acknowledgement

This study was partly supported by The Swedish Foundation for Strategic Research.

References

[1] M.A. Jobling, P. Gill, Encoding evidence: DNA in forensic analysis, Nat. Rev. Genet. 5 (2004) 739–751.
[2] D. Primorac, M.S. Schanfield, Application of forensic DNA testing in the legal system, Croat. Med. J. 41 (2000) 32–46.
[3] A.P. Dawid, J. Mortera, V.L. Pascali, Non-fatherhood or mutation? A probabilistic approach to parental exclusion in paternity testing, Forensic Sci. Int. 124 (2001) 55–61.
[4] T. Egeland, P. Mostad, Statistical genetics and genetical statistics: a forensic perspective, Scand. J. Stat. 29 (2002) 297–307.
[5] http://www.dna-view.com.
[6] I.W. Evett, B.S. Weir, Interpreting DNA Evidence, Statistical Genetics for Forensic Scientists Sinauer, Sunderland, MA, 1998.
[7] D.J. Balding, Weight-of-Evidence for Forensic DNA, John Wiley and Sons Ltd., Chichester, UK, 2005.
[8] National Research Council, Committee on DNA Forensic Science: An Update & Commission on DNA Forensic Science: An Update. The Evaluation of Forensic DNA Evidence, National Academy Press, Washington, DC, 1996.
[9] T. Egeland, P.F. Mostad, B. Mevag, M. Stenersen, Beyond traditional paternity and identification cases. Selecting the most probable pedigree, Forensic Sci. Int. 110 (2000) 47–59.
[10] E.M. Shepard, R.J. Herrera, Iranian STR variation at the fringes of biogeographical demarcation, Forensic Sci. Int. 158 (2006) 140–148.
[11] D. Podini, A. Nuccitelli, N. Vitale, L. Barbetta, S. Moscarelli, F. Fiorentino, Studio preliminare di 15 loci STR su di un campione di soggetti somali inserito in un programma internazionale di ricongiungimento familiare, in: Atti del XIX Congresso Nazionale Ge. F. I.-Genetisti Forensi Italiani, 2002.
[12] M. Regueiro, J.C. Carril, M.L. Pontes, M.F. Pinheiro, J.R. Luis, B. Caeiro, Allele distribution of 15 PCR-based loci in the Rwanda Tutsi population by multiplex amplification and capillary electrophoresis, Forensic Sci. Int. 143 (2004) 61–63.
[13] A.P. Dawid, J. Mortera, V.L. Pascali, D. Van Boxel, Probabilistic expert systems for forensic interference from genetic markers, Scand. J. Stat. 29 (2002) 577–596.
[14] http://www.cstl.nist.gov/div831/strbase/mutation.htm.
[15] B. Rannala, J.L. Mountain, Detecting immigration by using multi locus genotypes, Proc. Natl. Acad. Sci. U.S.A. 94 (1997) 9197–9201.
[16] C.H. Brenner, Some mathematical problems in the DNA identification of victims in the 2004 tsunami and similar mass fatalities, Forensic Sci. Int. 157 (2006) 172–180.
[17] T. Egeland, G. Storvik, A. Salas, Statistical considerations for haploid databases, submitted for publication.