Accurate assessment of the weight of evidence for DNA mixtures by integrating the likelihood ratio

Slooten, K.

published in
Forensic Science International: Genetics
2017

DOI (link to publisher)
10.1016/j.fsigen.2016.11.001

document version
Publisher's PDF, also known as Version of record

Forensic Science International: Genetics 27 (2017) 1–16

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

Research paper

Accurate assessment of the weight of evidence for DNA mixtures by integrating the likelihood ratio

Klaas Slooten [a,b,*]

[a] Netherlands Forensic Institute, P.O. Box 24044, 2490 AA The Hague, The Netherlands
[b] VU University Amsterdam, De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 18 May 2016
Received in revised form 26 October 2016
Accepted 6 November 2016
Available online 16 November 2016

Keywords:
Likelihood ratios
DNA mixtures
Semi-continuous model
Allelic dropout
Low template DNA
Deconvolution
Weight of evidence

ABSTRACT

Several methods exist for weight of evidence calculations on DNA mixtures. Especially if dropout is a possibility, it may be difficult to estimate mixture specific parameters needed for the evaluation. For semi-continuous models, the LR for a person to have contributed to a mixture depends on the specified number of contributors and the probability of dropout for each. We show here that, for the semi-continuous model that we consider, the weight of evidence can be accurately obtained by applying the standard statistical technique of integrating the likelihood ratio against the parameter likelihoods obtained from the mixture data. This method takes into account all likelihood ratios belonging to every choice of parameters, but LR's belonging to parameters that provide a better explanation to the mixture data put in more weight into the final result. We therefore avoid having to estimate the number of contributors or their probabilities of dropout, and let the whole evaluation depend on the mixture data and the allele frequencies, which is a practical advantage as well as a gain in objectivity. Using simulated mixtures, we compare the LR obtained in this way with the best informed LR, i.e., the LR using the parameters that were used to generate the data, and show that results obtained by integration of the LR approximate closely these ideal values. We investigate both contributors and non-contributors for mixtures with various numbers of contributors. For contributors we always obtain a result close to the best informed LR whereas non-contributors are excluded more strongly if a smaller dropout probability is imposed for them. The results therefore naturally lead us to reconsider what we mean by a contributor, or by the number of contributors.

1. Introduction

Several methods exist for weight of evidence calculations on DNA mixtures where dropout is a possibility. Broadly speaking, there are two types of such methods: the semi-continuous models that take into account the set of recorded alleles of the mixture and calculate a mixture likelihood using as parameters the number of contributors to the mixtures and parameters that describe dropout and possibly also drop-in probabilities; and continuous models that take also the peak heights into account and need a more refined probabilistic model that allow to compute a likelihood for the observed peak heights, also needing as input parameter the number of contributors. In this article we will focus on the semi-continuous models. Several of these exist, e.g. LRmixStudio (www.lrmixstudio.org, see also [1]), LabRetriever (cf. [2]), LikeLTD (cf. [3]), and our own implementation MixKin (cf. [4]). In practice it may be difficult to assess the number of contributors and the probabilities of dropout. The latter are sometimes (e.g., for LRmixStudio) chosen by the user and can then be varied to carry out a sensitivity analysis, or they can be estimated by the software (as is the case for LikeLTD). It is therefore logical that estimating and modelling the probability of dropout has received considerable attention in the literature, cf. [5–7], and so has the question how well the number of contributors to a mixture can be estimated (cf. [8–11]).

On the other hand, the notion of number of contributors is perhaps less useful in case the contributors do not have all their alleles recorded in the mixture. It should intuitively not make much difference whether or not an extra person with a high degree of dropout is present or not, as long as this is not our contributor of interest targeted by LR calculations. In this article, we will therefore take a different approach: we obtain the weight of evidence by integrating away the dropout probabilities. This means that no attempt is made to estimate them and compute a LR

* Correspondence to: Netherlands Forensic Institute, P.O. Box 24044, 2490 AA The Hague, The Netherlands.

http://dx.doi.org/10.1016/j.fsigen.2016.11.001
1872-4973/

based on these estimates. At the same time, we also refrain from having to specify the exact number of contributors: instead we take the maximal number of contributors and let the probabilities of dropout for all of them vary between zero and one. If we want we can also restrict them to a range within which they should reasonably lie in view of the mixture data. Integration over nuisance parameters is a standard statistical technique also employed by other mixture software (e.g., STRmix, [12]), an alternative being maximum likelihood estimation (e.g., [13,14]).

To understand intuitively why this can be a viable approach, let us first go back to a more simple situation, that of a single source trace which is not affected by dropout or dropin. A human interpreter, when seeing such a profile, will typically conclude that this is indeed a single source profile. The interpreter will do so, because it is extremely unlikely for a mixture of two or more persons to give rise to a mixture profile that could have been a single source profile, i.e., with one or two alleles observed on each locus. In case replicate analyses are performed which give rise to the exact same profile, this strengthens the conclusion further. The reason why an interpreter makes this decision essentially is based on likelihoods: it is so extremely unlikely for a mixture to consistently produce a profile that could have also been produced by one and the same single contributor, that these possibilities are discarded by the interpreter and single source is concluded.

The structure of this paper is as follows. In Section 2 we will begin by giving some more details on the semi-continuous model and our implementation of it in the symbolic programming language Mathematica, before explaining the method of integrating over the set of possible dropout probabilities.

Before we carry out systematic calculations, in Section 3 we give an example of a two-person mixture, to further illustrate the ideas behind integration rather than estimation of the needed parameters.

After that, in Section 4 we return to the single source profile situation sketched above, to see that indeed the weight of evidence is hardly affected by assuming the possibility that there might be more than one contributor.

Then we study some types of mixtures systematically. In all cases, we consider mixtures where none of the donors are supposed to be known, and we evaluate the LR for a person of interest to be a donor, where this person is in reality either a true donor or a non-donor. We do so for concision and because we believe that mixtures without known donors are harder to evaluate than those where some donors are known. In Section 5 we investigate various types of two-person mixtures: mixtures with a clear major/minor donor, and mixtures where both donors have contributed equally. We compare also the results obtained if we treat these mixtures as two-person mixtures or three-person mixtures. In Section 6 we look at three-person mixtures and carry out similar calculations.

Until this point, the purpose of the paper can be summarized as showing that, if the semi-continuous model is applicable to mixture data, then the method of integrating away the dropout parameters is an accurate way of retrieving the weight of evidence that we would have obtained with the correct dropout probabilities. In order to show this, it is inevitable to use simulated rather than lab-generated mixtures, since only for simulated data can we know the actual probabilities of dropout. So this shows that, if the model is applicable, then it is reliable. For the model to be applicable, it is essential that the modelling assumptions are met. In particular, dropout probabilities should not vary (substantially) across replicates and therefore these replicates should have been obtained under identical conditions. For instance, the model reported upon here supposes that the probabilities of dropout are constant over the loci, for each contributor. It is however easy to incorporate degradation (increasing dropout probabilities with increasing fragment size) and in Section 7 we briefly look into the effect of ignoring degradation, by comparing the obtained results on mixtures affected by degradation when we do, or do not, take it into account. We also look into another issue, the dropout probability for homozygotes as compared to that for heterozygotes.

Finally we summarize our findings and conclusion in the discussion.

2. Methods

2.1. Semi-continuous model

We start by recalling the characteristics of the semi-continuous model that we use in this article. This description is a summary of the one given in [15] and we refer the reader to that paper for further details. Suppose that a mixture has n contributors. Let their dropout probabilities be $d_1, \ldots, d_n$ with $0 \le d_i \le 1$, and let $c \ge 0$ be the parameter describing the expected number of alleles dropping in per locus. We define the probability that allele a is detected in mixture M as

$$P_{\vec{d},c}(a \in \mathcal{M}|\vec{g}) = 1 - e^{-cp_a} \prod_{i=1}^{n} d_i^{n_{i,a}}, \tag{2.1}$$

where $n_{i,a} \in \{0, 1, 2\}$ is the number of alleles a present in g_i, the genotype of contributor i (by definition, $0^0 = 1$) and p_a is the allele frequency of allele a. Note that, when $c = 0$, we see from this formula that an allele is recorded unless it drops out for all the contributors that have that allele. In [4,15] we have used the approximation $e^{-cp_a} \approx 1 - cp_a$ for $c \ll 1$. To compute the probability that the observed mixture $\mathcal{M}$ is equal to the set of alleles M, one simply uses (2.1) to obtain

$$P_{\vec{d},c}(\mathcal{M} = M|\vec{g}) = \prod_{x \in M} P_{\vec{d},c}(x \in \mathcal{M}|\vec{g}) \prod_{x \notin M} P_{\vec{d},c}(x \notin \mathcal{M}|\vec{g}). \tag{2.2}$$

We also note that contaminant alleles may coincide with alleles of donors, and thus do not necessarily lead to an allele in the mixture that is not present in any of the contributor's genotypes. In particular they may undo the effect of allelic dropout. For the interpretation of this model, notice that $1 - d_i$ is the fraction of alleles of contributor i that we would expect to observe if the other contributors would not have been there and in the absence of drop-in. The probability of dropout d_i is therefore a prospective probability, allowing to attach likelihoods to mixture data, rather than a retrospective probability making statements about the probability that alleles have been unrecorded in the mixture data. We refer to [16] for further discussion of the connection between these.

The parameter c is equal to the expected number of alleles dropping in per locus, assuming that the number of alleles dropping in is Poisson distributed with parameter c and that the alleles that drop in are sampled at random according to the allele frequencies.

The probability to observe $\mathcal{M} = M$ when some of the donors have unknown genotypes is obtained by summing (2.2) over the set of possible genotypes for these donors, weighted by their prior probability to be the donor's genotypes. Suppose that the expected population frequency of genotype g is denoted f_g, by which we mean the probability that a person chosen at random from the population has genotype g. In standard mixture calculations, without relatedness in the hypotheses, and without applying the θ-correction (cf. [17]), one sets

$$P_{\vec{d},c}(D_i = g) = f_g \tag{2.3}$$

as the a priori distribution of the genotype of D_i of donor i. Note that this is of course independent of $\vec{d}$ and c: regardless of the mixture or of how we evaluate it, we assume prior to having any mixture data that each donor is a random person from the population.

If there are several replicate profiles available from the same trace, then we will assume that the replicate profiles are conditionally independent of each other given the probabilities of dropout $\vec{d}$ and the drop-in parameter c. Thus, the probability to see all replicates is simply the product of the probabilities to see each of the replicates separately. We note however that the assumption that the parameters $\vec{d}$ and c are the same over all replicates, and the conditional independence, can only be justifiable if all replicates are the results of independent PCR's using the same technology (PCR cycle number, CE settings).

For a hypothesis H describing the contributors of the mixture, we let

$$P_{\vec{d},c}(M|H)$$

be the mixture likelihood. Now suppose that we have two hypotheses H_1 and H_2, and that the difference between them is that H_1 postulates a person of interest (PoI) with genotype g to be a mixture contributor, whereas H_2 has replaced that contributor by an unknown individual. There may be uncontested contributors, postulated both by H_1 and H_2. The likelihood ratio for the semi-continuous model with parameters $\vec{d}$ and c is then

$$\mathrm{LR}_{\vec{d},c}(M,g) = \frac{P_{\vec{d},c}(M|H_1)}{P_{\vec{d},c}(M|H_2)}. \tag{2.4}$$

2.2. Implementation

We have implemented the model just described in our script MixKin written in the programming language Mathematica (cf. [4]). It is possible to do numerical computations with this script, but also, since Mathematica is a symbolic programming language, to compute the mixture likelihood algebraically as a function of the variables $\vec{d}$ and c. The resulting value for a specific choice of parameters can be rapidly obtained from this expression which facilitates sensitivity analyses. Also, as we have used in this paper, the algebraic expression for the LR can be (numerically) integrated. For this we have used the built-in command NIntegrate with default settings.

The results in this paper have been obtained from simulations on the 15 autosomal NGM loci, using allele frequencies taken from [18].

2.3. Weight of evidence

Given mixture data, a LR calculation $\mathrm{LR}_{\vec{d},c}(M,g)$ needs a specification of the number of contributors and their probabilities of dropout $\vec{d}$, and of the drop-in parameter c. One approach to arrive at such an estimate is to consider the number of recorded alleles as an indication for the number of contributors, to estimate dropout probabilities from peak heights of recorded alleles, etc. One can use such approaches to arrive at plausible ranges for the dropout parameters. For example, for a certain mixture profile one may be sure that it has at least two contributors, a major one and a minor one, and possibly a third one that must have a high degree of dropout. One may for example estimate that the probability of dropout for the major donor, d_1, lies between 0 and 0.2 and that for the second donor we have d_2 between 0.4 and 0.6. Finally the third contributor, if present, has d_3 satisfying $0.8 \leq d_3 \leq 1$, where $d_3 = 1$ corresponds to the situation where there are only two contributors.

It is then a natural step to define the mixture likelihood as

$$P(M|H) = \int_{x_1=0}^{0.2}\int_{x_2=0.4}^{0.6}\int_{x_3=0.8}^{1} P_{(x_1,x_2,x_3),c}(M|H)dx_1dx_2dx_3.$$

Notice that this amounts to treating the dropout probabilities as nuisance variables with uniform distribution over their domain. In this formula, we have kept c fixed, so we have not integrated over some distribution for this parameter. In principle it would be possible to do so, but it is computationally cumbersome to perform another integration. Also, unless drop-in is needed to make the specified number of contributors possible, the LR often does not strongly depend on it for small (realistic) values of c. Finally one may argue that whereas the probabilities of dropout are trace dependent, the probability of alleles dropping in is less so and can be estimated by using negative controls.

If we want to calculate the LR for a person being a contributor to the mixture, we need to specify which contributor it is: the one with dropout between 0 and 0.2, between 0.4 and 0.6 or between 0.8 and 1.

Now, we may even argue that we may as well extend the integration for all three dropout variables d_i to [0, 1], since (if our estimates are accurate) the mixture likelihood is going to be small over the part of $[0, 1] \times [0, 1] \times [0, 1]$ that is outside the region $[0, 0.2] \times [0.4, 0.6] \times [0.8, 1]$ and its permutations. Therefore, we will in this article start with a uniform prior over all dropout probabilities $(d_1, d_2, \ldots, d_n) \in [0, 1]^n$, and we define the LR as

$$\mathrm{LR}(M,g) = \frac{\int_R P_{\vec{x},c}(M|H_1)d\vec{x}}{\int_R P_{\vec{x},c}(M|H_2)d\vec{x}}, \tag{2.5}$$

where $R = [0, 1]^n$ and $\vec{x} = (x_1, \ldots, x_n)$. We do not integrate over c for the abovementioned reasons. Note that we may also write

$$\mathrm{LR}(M,g) = \frac{\int_R \mathrm{LR}_{\vec{x},c}(M,g)P_{\vec{x},c}(M|H_2)d\vec{x}}{\int_R P_{\vec{x},c}(M|H_2)d\vec{x}}. \tag{2.6}$$

This provides the likelihood ratio for the hypotheses H_p: the mixture has at most n donors and, in addition to the uncontested contributors, the person of interest with genotype g is a contributor, versus H_d: the mixture has at most n donors and all donors apart from the uncontested contributors are unknown.

Contrary to the first example where we specified ranges of the dropout probabilities for each contributor, we now no longer can distinguish between the contributors. This approach should nonetheless give good results if the mixture likelihoods $P_{\vec{d},c}(M|H_p)$ are concentrated mainly around the actual dropout probabilities.

2.4. Interpretation

Notice that we can rewrite (2.6) as

$$\mathrm{LR}(M,g) = \int_R \mathrm{LR}_{\vec{x},c}(M,g)\pi_{M,H_2}(\vec{x},c)d\vec{x}, \tag{2.7}$$

where

$$\pi_{M,H_2}(\vec{d},c) = \frac{P_{\vec{d},c}(M|H_2)}{\int_R P_{\vec{x},c}(M|H_2)d\vec{x}}, \tag{2.8}$$

which corresponds to the probability density function of the dropout probabilities, taking into account the mixture data and (if any) the uncontested contributors postulated by H_2. Thus, $\mathrm{LR}(M, g)$ is a weighted average of the $\mathrm{LR}_{\vec{d},c}(M,g)$ for all parameters $\vec{d}$, weighted by the probability that these are the correct probabilities of dropout given the mixture data. Therefore, $\mathrm{LR}(M, g)$ will be largely determined by the values of $\mathrm{LR}_{\vec{d},c}(M,g)$ for plausible values of the parameters. This also means that, if we restrict the integration domain $R = [0, 1]^n$ to a region $R' \subset R$ that is deemed to contain all plausible values, this may have some computational advantage (since we carry out numerical rather than algebraic integration), but is not going to alter the obtained $\mathrm{LR}(M, g)$ significantly if the domain R' is well chosen and does not miss

plausible values of the dropout probabilities for which $LR_{\vec{d},c}(M,g)$ differs substantially from the values it takes inside R'.

Notice also that, whereas $LR_{\vec{d},c}(M,g)$ factorizes over loci if these loci are independent, this is no longer the case for the integrated LR(M, g).

2.5. Best informed LR

In the sequel of this paper, we will compare the LR obtained with the generating parameters $\vec{d}$ and c, with LR(M, g) obtained by integration. We will call the LR obtained with the true probabilities of dropout (i.e., the values that generated the mixtures) the *best informed* LR. Indeed this is the LR that we would calculate if our knowledge about the mixture generation parameters had been complete. Less knowledge about these parameters should intuitively correspond to a worse resolution between donors and non-donors, meaning – on average – a lower weight of evidence for actual donors and a softer exclusion of non-donors. However, this of course does not mean that for a particular comparison of a mixture and a donor, the largest LR is obtained with the most accurate probabilities of dropout. The best informed LR is the value that we would ideally report, since it reflects the evaluation of the data according to the model that actually generated the data. In actual casework, it is of course difficult to estimate these parameters, and therefore we investigate in this paper how close the LR obtained by integration is to the best informed LR. We now need some statistic to measure the distance between the best informed LR and a differently obtained one. In order to do so, we consider the weight of evidence (WoE), defined as Log_{10}(LR) as the main quantity of interest. Indeed, contrary to the LR, the weight of evidence may be interpreted as an amount of information, measured in bans, the base-10 equivalent of the (binary) bit (cf. [19]). Comparing the best informed weight of evidence with a differently obtained one, allows to measure how much information has been lost, or perhaps to which extent we overestimate the information contained in the mixture about its contributors.

Note that, if the dropout probabilities are $d_1, \ldots, d_n$, then there are in principle n best informed LR's for a person of interest, since we may compute a LR trying to identify the person of interest with the donor with probability of dropout equal to d_i for $1 \leq i \leq n$. For an actual donor, we will only consider the best informed LR to be the one corresponding to the probability of dropout of that true donor.

As relevant statistic we will compare the weight of evidence Log_{10}(LR(M, g)) with the best informed weight of evidence, the base-10 logarithm of the best informed LR. We will mostly consider the ratio $Log_{10}(\mathrm{LR}(M,g))/Log_{10}(\mathrm{LR}_{\vec{d},c}(M,g))$, which we can interpret as corresponding to the amount of information obtained by integration, relative to the amount of information that is really present.

We have used here a uniform prior probability density function over all the probabilities of dropout. It is of course necessary to specify a prior if we wish to obtain a posterior probability density function $\pi_{M,H_2}(\vec{d})$ for these parameters, which is based on the mixture data. Whether or not the resulting method is able to generate likelihood ratios that come close to the best informed LR, will be determined by a combination of the choice of the prior, and its relevance: it is of course preferable to be in a situation with sufficient data so that the results will be about the same for all reasonable choices of priors. Whether or not it is desirable to carry out such an integration is a source of debate (cf. [20,21]). In the research described here, we will show that with this particular choice of prior we calculate a LR that is very close to the best informed LR, and that therefore the situation here lends itself very well to the approach of integrating out the nuisance parameters of dropout. However, we also note that the quality of LR(M, g) will obviously depend on the number of loci and replicates. In this article, we have worked with the NGM loci and considered three replicates per mixture.

2.6. Deconvolution

As we have shown in [16], for fixed parameters $\vec{d}, c$, LR calculations are equivalent to mixture deconvolution. Indeed, suppose that we have a suspect with genotype g, and H_p identifies this person with donor D_1, whereas H_d has an unrelated unknown individual as donor instead of the suspect. Then, recalling that f_g denotes the population frequency of profile g,

$$f_g \mathrm{LR}_{\vec{d},c}(M,g) = P_{\vec{d},c}(D_1 = g | M, H_1). \tag{2.9}$$

When we plug this into (2.6) we get its analogue:

$$f_g \mathrm{LR}(M,g) = \frac{\int_R P_{\vec{x},c}(D_1 = g|M,H_1) P_{\vec{x},c}(M|H_2) d\vec{x}}{\int_R P_{\vec{x},c}(M|H_2) d\vec{x}} = \int_R P_{\vec{x},c}(D_1 = g|M,H_1) \pi_{M,H_2}(\vec{x},c) d\vec{x} = P(D_1 = g|M,H_1).$$

Thus, we see that also for the LR obtained by integration, we still have the equivalence between likelihood ratio calculation and deconvolution that holds for specific choices $\vec{d}, c$. Similar to LR(M, g), the probability $P(D_1 = g \mid M, H_1)$ is now obtained as a weighted average of this probability for all parameter values, where the weights are given by the probability density function for the parameters given the mixture data.

3. Example

We start with an example that illustrates and motivates the research presented in this paper. We consider a randomly generated mixture with two contributors S_1 and S_2, the first with probability of dropout $d_1 = 0.1$, and the other with $d_2 = 0.5$, and we generate three replicates using these parameters (we set the drop-in parameter at 0). If we calculate the LR using the parameters that have been used to generate the mixture, we get as best informed LR's the values $\mathrm{LR}_{(0.1,0.5),0}(M, g) = 10^{14.70}$ for the major donor and $\mathrm{LR}_{(0.1,0.5),0}(M, g) = 10^{9.033}$ for the minor donor. In reality of course, we do not know the parameters of the semi-continuous model that correspond to casework mixture data, and we therefore first investigate whether the mixture likelihoods can still to some extent reveal them if we integrate the mixture likelihoods over the parameter space.

We have compared the best informed LR for the major and minor contributor, with the LR's obtained by integration where we treat the mixture either as a trace with at most 2, or at most three, contributors. For each number of contributors, we have carried out two integrations: the integration where all dropout variables d_i range between 0 and 1; and integration of the likelihood function where the dropout parameters are taken to be equal. Thus, for the three-person case this is the integral $\int_0^1 P_{(x,x,x),0}(M|H)dx$. The resulting weights of evidence are presented in Table 1.

We see from this table that, when we make no attempt to estimate the probabilities of dropout, we do not lose a lot of information. Even when we treat this two-person mixture simply as a trace with at most three contributors and without any attempt to estimate or restrict dropout probabilities, we still lose only one ban of weight of evidence. For the major donor, this corresponds to a WoE that is 13.70/14.70 = 93% of what we would have had in the best informed position, and for the minor donor it corresponds to 88%. Notice also that in this case where the probabilities of dropout are in reality quite different from each other, we lose a lot of weight of evidence by integrating if we assume the probabilities of

Table 1
Weight of evidence obtained with various approaches.

Donor	Obtained Log_{10} (LR)				
	Best informed	Integration (max 2 persons)		Integration (max 3 persons)	
		Unequal dropout	Equal dropout	Unequal dropout	Equal dropout
Major	14.70	13.99	11.62	13.70	11.04
Minor	9.033	8.275	4.230	7.969	5.185

dropout of all donors to be the same. For the major donor, we preserve about 75% of the weight of evidence, but for the minor donor only about half.

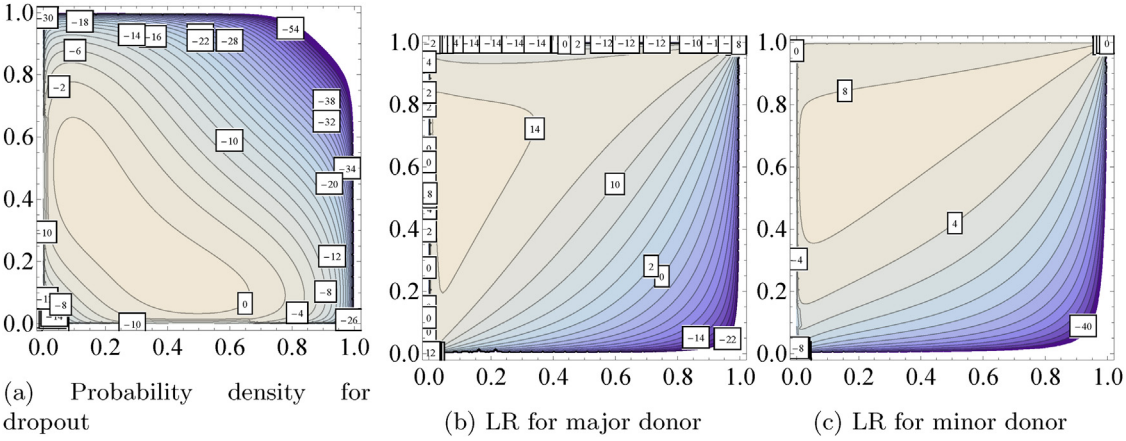
In order to illustrate how one arrives at these values for LR(M, g) we plot in Fig. 1 the terms of (2.7): in Fig. 1a we plot the probability density function $\pi_{M,H_2}(\vec{d}, 0)$ (on log-scale), and in Fig. 1b and c we plot $\text{LR}_{\vec{d},0}(M, g)$ for the major and minor donor as a function of d_1 and d_2. We see that indeed the integrated LR(M, g) is determined mainly by the parameters (d_1, d_2) for which the likelihood $\pi_{M,H_2}((d_1, d_2), 0)$ and the likelihood ratios $\text{LR}_{(d_1,d_2),0}(M, g)$ for these parameters are maximal, i.e., when we have evidence both in favor of the dropout parameters being the ones that describe the mixture data and in favor of the person of interest having contributed.

4. Single source traces

To further illustrate the principle, we now turn to single source traces unaffected by dropout or drop-in where we again compare the best informed LR with the LR obtained by integration assuming a maximal number of one, two or three contributors. In case we test the actual donor of the trace, the best informed LR is the inverse of the expected population frequency f_g of the profile g at hand. Indeed, the best informed LR is obtained when we know that this is a single source profile, so the LR obtained for a person of interest is either zero (for a PoI with a different genotype) or $1/f_g$, if the PoI has genotype g. For these single source traces, we have assumed $c = 0$ both for the generation of the trace profiles and for the LR calculations.

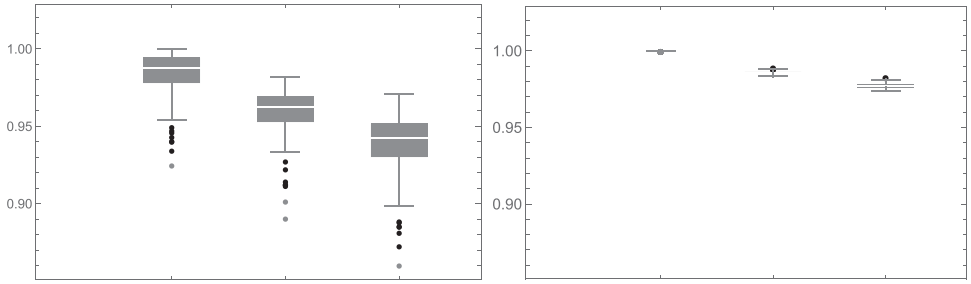
4.1. Actual donor

For an actual donor we will obtain a full match between the trace profile and the genotype of the PoI. We will compare the best informed LR by the LR that we obtain by integration when we suppose the trace is a mixture of at most n persons where $n \in \{1, 2, 3\}$. We then take as statistic the amount of preserved weight of evidence, i.e., $Log_{10}(\text{integratedLR})/Log_{10}(\text{bestinformedLR})$. We provide summaries of this statistic in Fig. 2, based on 100 analyses. In this figure, as well as in the figures that follow, we have presented the results in the form of box-plots. The boxes ranges from the 25% quartile (Q1) to the 75% quartile (Q3) with a bar indicating the median. The whiskers extend to the last point that is within 1.5 times the inter-quartile range (Q3–Q1) from the box, and remaining points are plotted separately as outliers.



(a) Probability density for dropout (b) LR for major donor (c) LR for minor donor

Fig. 1. Parameter likelihoods and LR's for both donors (all on Log-scale), as a function of the probabilities of dropout, for the mixture discussed in the Example.



(a) Preserved WoE by integration resp. as one, two, three-person trace based on one replicate (b) Preserved WoE by integration resp. as one, two, three-person trace based on three identical replicates

Fig. 2. Preserved weight of evidence if LR is obtained by integration of likelihoods, for single source traces without dropout.

From this figure we see that hardly any information is lost if we, rather than treating the trace with the correct parameters, treat it as a trace that may have come from at most three contributors, where all dropout probabilities are a priori equally likely. Considering it as a trace of at most three persons leads to a preservation of weight of evidence of on average about 95% for these profiles if we have one replicate. If we suppose that the trace has been analysed three times, each time with the same result, then the amount of preservation is even higher as we can see in Fig. 2b. If we treat it as a one person trace but integrate over the dropout probability for its donor, we do not lose any information any more, but even when we admit up to three contributors we still preserve more than 97,5% of the weight of evidence, on average. This result is understandable using the logic that was described in the introduction of this paper: the probability density $\pi_{M,H_2}(\vec{d})$ is so small outside the regions corresponding to a single contributor without dropout, that LR(M, g) is almost completely determined by its values at these parameters: even if we assume a priori (where a priori here means prior to having any data from the trace) that we are going to analyze a trace left by possibly three individuals, the trace data indicate a single contributor so strongly that our prior assumptions on the number of contributors or their dropout probabilities have become almost completely irrelevant.

5. Two-person mixtures

We now turn to mixtures and start to consider two-person mixtures. We look at mixtures with a major/minor donor (for which we take $d_1 = 0.1$, $d_2 = 0.5$) and mixtures where both donors have contributed equally (where we let $d_1 = d_2 = 0.3$). For both types of mixtures we have, as for the single source traces, generated 100 mixtures with these parameters, supposing that we have three replicates for each mixture. Then, for each such simulated mixture, we have compared the best informed LR (calculated with the parameters that generated the data) with several methods to obtain the weight of evidence: the LR obtained by integration, either unrestricted of restricted, assuming two or three possible donors, and the point estimates $\mathrm{LR}_{\vec{d}}(M, g)$ corresponding to various choices. Both for the generation of the mixtures as well as for the LR calculations we have used $c = 0.05$.

5.1. Mixtures with a major and a minor donor

We first consider mixtures with a major donor ($d_1 = 0.1$), a minor donor ($d_2 = 0.5$), of which we have three replicates on the NGM loci.

5.1.1. Major donors

We first consider the actual major donors. To evaluate whether obtaining a LR by integration is advantageous over estimating the probabilities of dropout, we compare the best informed LR to the LR obtained with various techniques: unrestricted integration assuming at most two contributors, restricted integration assuming two contributors with $0 \leq d_1 \leq 0.2$, $0.4 \leq d_2 \leq 0.6$, unrestricted integration assuming at most three contributors, restricted integration assuming three contributors with $0 \leq d_1 \leq 0.2$, $0.4 \leq d_2 \leq 0.6$, $0.8 \leq d_3 \leq 1$, and the point estimates $\mathrm{LR}_{\vec{d},c}(M, g)$ for $d_1 = d_2 \in \{0.1, 0.3, 0.5\}$. For all these methods, we have compared Log_{10}(calculated LR)/Log_{10}(bestinformed LR). The results are presented in Fig. 3.

We conclude from Fig. 3 that the LR obtained by integration, even if three persons are assumed, performs much better than evaluating the LR assuming specific probabilities of dropout that are equal for both contributors. In other words, we obtain a weight of evidence that is closer to the best informed weight of evidence if we make no assumptions at all other than that the trace has at most three contributors, than if we evaluate the trace with the correct number of contributors, but using the same probability of dropout for both. We notice also that restricted integration performs slightly better than unrestricted integration but not to a great extent: as we already mentioned in Section 2.4, this is because the values that are discarded in the restricted integration, are among the least plausible given the mixture data and therefore their contribution to the LR obtained by unrestricted integration is small.

5.1.2. Minor donors

We now consider the actual minor donors. Again we compare the best informed LR to the LR obtained with various techniques: unrestricted integration assuming at most two contributors, restricted integration assuming two contributors with $0 \leq d_1 \leq 0.2$, $0.4 \leq d_2 \leq 0.6$, unrestricted integration assuming at

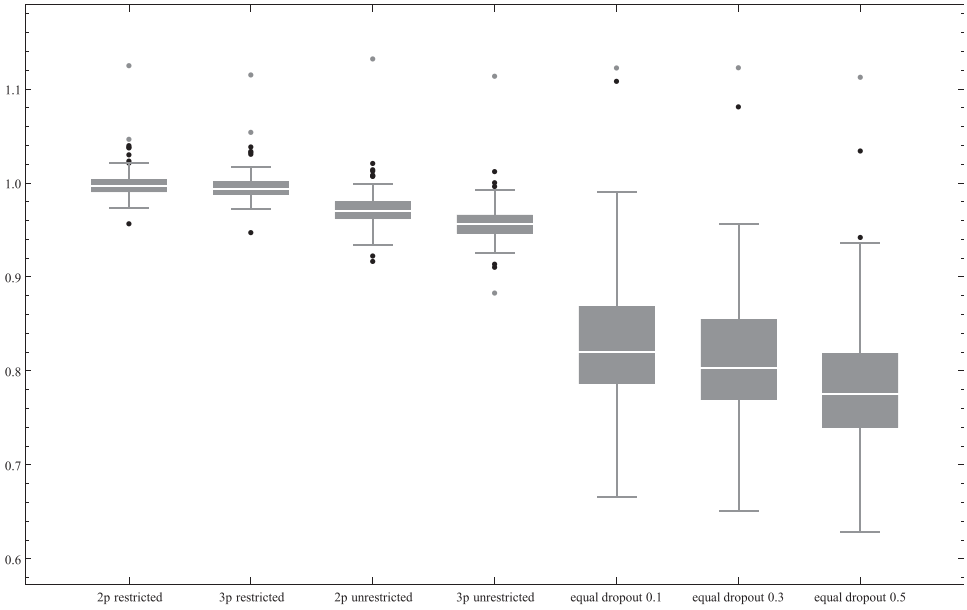


Fig. 3. Preserved fraction of weight of evidence for major donors in two-person mixtures with probabilities of dropout (0.1, 0.5), when LR's are calculated using various methods (detailed in text).

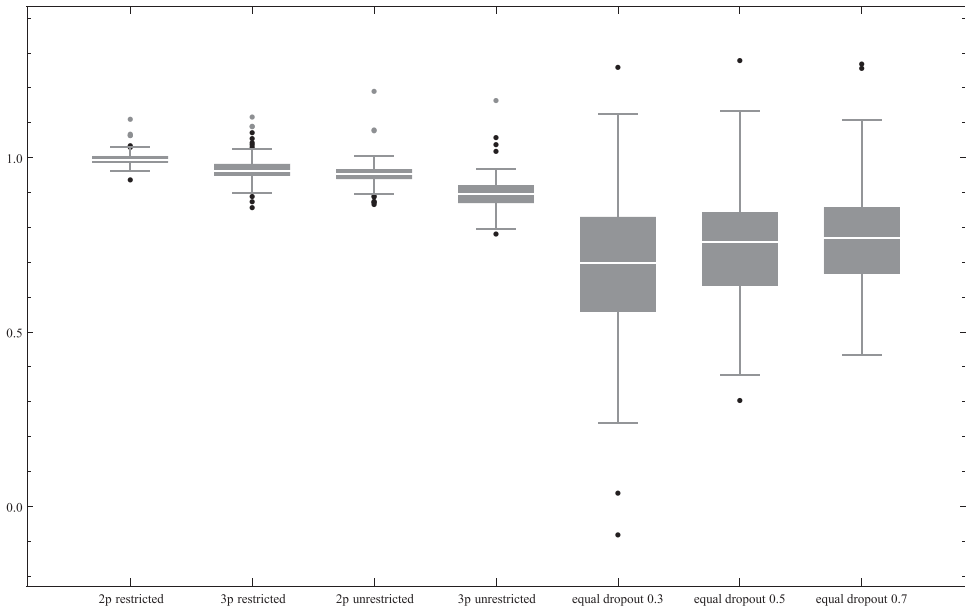


Fig. 4. Preserved fraction of weight of evidence for minor donors in two-person mixtures with probabilities of dropout (0.1, 0.5), when LR's are calculated using various methods.

most three contributors, restricted integration assuming three contributors with $0 \leq d_1 \leq 0.2, 0.4 \leq d_2 \leq 0.6, 0.8 \leq d_3 \leq 1$, and the point estimates $LR_{\vec{d},c}(M,g)$ for $d_1 = d_2 \in \{0.3, 0.5, 0.7\}$. For all these methods, we have compared Log_{10}(calculated LR)/Log_{10}(bestinformed LR). The results are presented in Fig. 4.

As we did for the major donors, we again conclude from Fig. 4 that the LR obtained by integration, even if three persons are assumed, performs much better than evaluating the LR assuming specific probabilities of dropout that are equal for both contributors. In fact, if we compare the results to those of the major donor, we see that especially the weight of evidence obtained with point values for the probabilities of dropout, is much more spread out compared to the best informed LR, whereas the weight of evidence obtained by integration, either restricted or unrestricted and regardless of assuming at most two or at most three contributors, is still close to the best informed values.

5.1.3. Non-contributors

For non-contributors, we need first to define what we now mean by the best informed LR. Indeed, the non-contributor in reality corresponds to none of the n donors, and therefore any LR obtained with the actual probabilities of dropout, identifying the non-contributor with donor i where $1 \leq i \leq n$ can be considered to be a best informed LR. Unless the probabilities of dropout are the same for all contributors, there is thus no single best informed LR to be used for comparison.

We therefore calculate the weight of evidence $Log_{10}(LR(M, g))$ for non-contributors in various ways: we compute the two best informed LR's, testing whether the PoI can be each of the contributors with $d_1 = 0.1$ or $d_2 = 0.5$, the LR calculated with $d_1 = d_2 = 0.3$, unrestricted integration assuming either two or three contributors, and restricted integration assuming three contributors with $0 \leq d_1 \leq 0.2,\ 0.3 \leq d_2 \leq 0.7,\ 0.7 \leq d_3 \leq 1$ where the LR

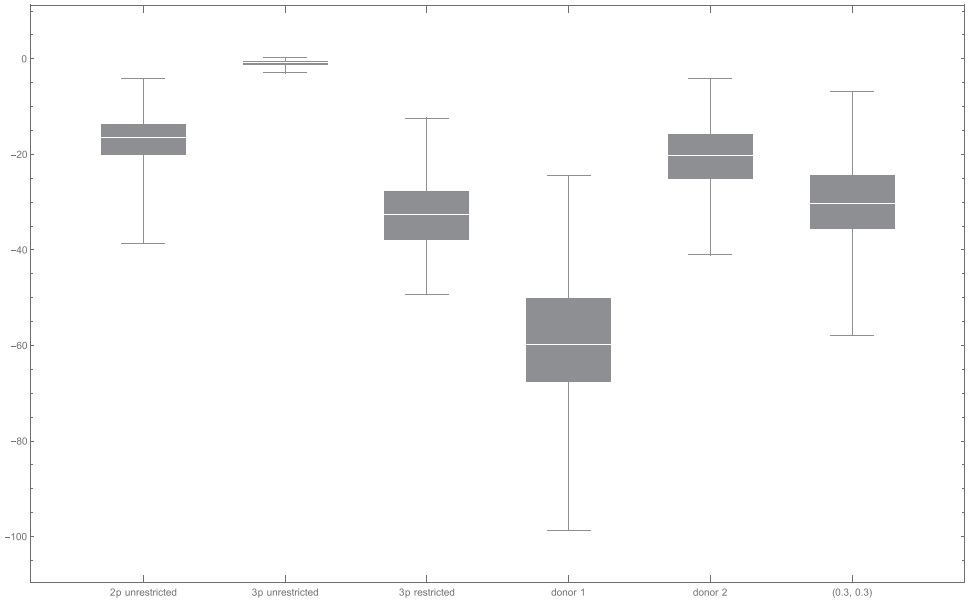


Fig. 5. Obtained weight of evidence by various methods, for non-contributors tested as possible contributors to two-person mixtures generated with $(d_1, d_2) = (0.1, 0.5)$.

tests if the PoI can be the first contributor. The results of these calculations can be found in Fig. 5.

We now see a marked difference: the LR's obtained for three possible contributors (the column "3p unrestricted" in Fig. 5 are hardly informative, whereas when assuming two possible contributors (the column "2p unrestricted") we obtain LR(M, g) $\ll 1$. We can understand this if we reason heuristically with likelihoods. If we assume that the mixture has only two contributors, then both contributors have a significant presence since their probabilities of dropout are such that many of their alleles are recorded, and the mixture likelihood will be maximal for two contributors which both do not have a very high probability of dropout (an example of this was given in Fig. 1a). None of these donors look like our random non-contributing PoI so LR(M, g) will be very small. That is, assuming two contributors, we cannot completely deconvolve the mixture, but we clearly have a strong counter-indication for the presence of our non-contributing person of interest. However, if we assume that there could be three contributors, then we can again reason with likelihoods: a two-person mixture, if treated as a trace coming from three persons, will have highest likelihoods for three persons in the situation where not all three contributors have a small probability of dropout. Therefore, for LR(M, g) the situation with a third donor with small contribution and correspondingly a high probability of dropout will dominate the result. Our person of interest does not look like a contributor with low dropout probability, so for LR(M, g) the comparison with the third contributor with high dropout will dominate. Such a comparison cannot result in a very decisive LR. What we arrive at is therefore very logical: it is hard to exclude that the person of interest has had a contribution to the mixture, but we can say that if he had then it must be a very small one, close to undetectable; which corresponds to a very high degree of dropout. In terms of a verbal description of the weight of evidence, this corresponds to a statement along the lines of not being able to exclude the person of interest as a donor, but no support in favour of this possibility has been found either: it can only be said that a contribution, if present, has to be minimal.

A way to specifically target a person with a non-negligible contribution to the mixture is to restrict the integration domain of the dropout probability for the targeted contributor. For example, we may integrate d_1, the probability of dropout for the person of interest, only between 0 and 0.5. For the other donors, we can let the d_i for vary between zero and one as before or also choose bounds for them. In that case, we specifically investigate whether there is support for the person of interest being a mixture donor with probability of dropout between 0 and 0.5. We call this restricted integration. To exemplify this, we have used restricted integration supposing that the first (major) donor has $0 \leq d_1 \leq 0.2$, the second (minor) donor has $0.3 \leq d_2 \leq 0.7$ and the third donor has $0.7 \leq d_3 \leq 1$. The results are given in the column "3p restricted" of Fig. 5. If we then calculate LR(M, g) for non-contributors testing whether they can be the first donor, we obtain much stronger evidence that the non-contributor indeed did not contribute.

5.1.4. Determining the number of donors

Finally we use this method to determine whether we can estimate the number of contributors to the above mixtures. In reality there are two: one with probability of dropout $d_1 = 0.1$ and the other with $d_2 = 0.5$. First, we need to realize what it means to be a contributor if we allow for dropout. Indeed, if we do not make any restrictions on the probability of dropout then a contributor need not have been seen in the replicates that we have obtained. In other words, if we conclude that there are two contributors then we exclude the possibility that alleles of another person will show up in future replicates, no matter how many of them we obtain. It will be much harder to conclude that this must be the case, than to conclude for example that there are only two contributors with a reasonable representation in the mixture. A second property of the integration method presented in this paper is that it uses a prior distribution over the probabilities of dropout. In this case, we have used a uniform prior, making all values a priori equally likely. In expectation, prior to having data, the probability of dropout for each donor is therefore 0.5 and these probabilities are subsequently updated using the mixture data to arrive at a posterior probability density function $\pi_{M,H_2}(\vec{d}, c)$ (cf. (2.8)). In view of these considerations, we have calculated for each mixture the weight of evidence in favour of the hypothesis that the mixture has two donors, versus having three donors, in two ways: first by unrestricted integration letting the probability of dropout range between 0 and 1 for all donors, and second by restricted integration letting these probabilities range between 0 and 0.7 for all donors. The last choice corresponds to donors with a sufficient contribution (here, arbitrarily set at a maximum probability of dropout of 0.7), whereas in the first choice we also consider donors with a smaller contribution. The results are presented in Fig. 6.

From these results we see that, when we make no restrictions on the probabilities of dropout, there is only very weak evidence in favour of having two contributors. Form the reasoning above this is understandable: we have obtained three replicates and these are generally different, from which we conclude that it is certainly not impossible that future replicates will show alleles that have not been seen in the ones that we have. Even though the currently obtained replicates are compatible with the mixture having two donors, it is therefore hard to say that we will never observe an allele that must come from a third person. However, if we restrict the probability of dropout to at most 0.7 for all donors, then we only consider possible contributors from which on average a fraction 0.3 or more of their alleles are detected per replicate. With this restriction on what it means to be a contributor, the evidence in favour of there being two rather than three contributors is much stronger.

5.2. Mixtures with two equally present contributors

For mixtures with two equally present contributors, there is no separation between major and minor donors any more. To see the extent to which this changes the results of our analyses, we have considered 100 two-person mixtures where both donors had the same probability of dropout $d_1 = d_2 = 0.3$. Again, we suppose that we have three replicates of each mixture and we have considered the NGM loci. We plot in Fig. 7 the preserved weight of evidence obtained by similar methods that we have used above: unrestricted integration assuming at most two contributors, restricted

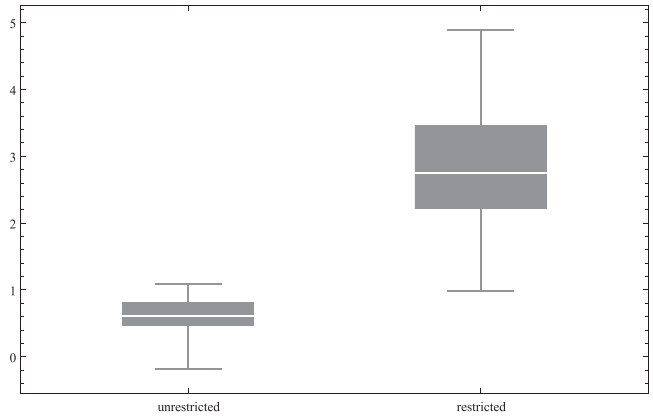


Fig. 6. Log_{10} (LR) for the hypotheses of the mixture having two versus three contributors.

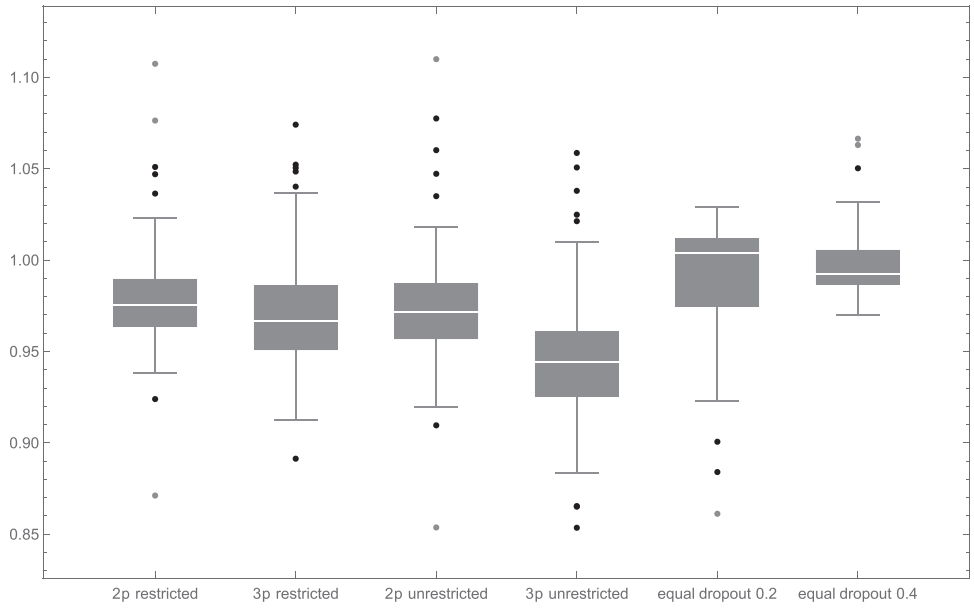


Fig. 7. Preserved fraction of weight of evidence for donors in two-person mixtures with probabilities of dropout (0.3, 0.3), when LR's are calculated using various methods.

integration assuming two contributors with $0 \leq d_1 \leq 0.5$, $0 \leq d_2 \leq 0.5$, unrestricted integration assuming at most three contributors, restricted integration assuming three contributors with $0 \leq d_1 \leq 0.5$, $0 \leq d_2 \leq 0.5$, $0.8 \leq d_3 \leq 1$, and the point estimates $LR_{x,x}(M, g)$ for $x \in \{0.2, 0.4\}$. For all these methods, we have compared Log_{10}(calculated LR)/Log_{10}(bestinformed LR). The results are presented in Fig. 7.

The results are now much closer to each other than in Figs. 3 and 4: all methods exhibit good performance. This is mainly due to the fact that the LR's obtained at the choices $d_1 = d_2 = 0.2$ and $d_1 = d_2 = 0.4$ are much closer to the best informed LR, than we saw for the case above where the contributors have different probabilities of dropout.

For non-contributors, we obtain essentially the same results (data not shown) as for the previously considered mixtures: if we calculate the integrated LR with at most three contributors, then we typically get a modest weight of evidence against contribution. If we either restrict the integration for the contributor of interest, or assume there are at most two contributors, then we get $LR(M, g) \ll 1$.

6. Three-person mixtures

Finally we consider three-person mixtures, where we have chosen $(d_1, d_2, d_3) = (0.1, 0.4, 0.7)$ and $c = 0$ for the data generation, and for the calculation of LR's for actual contributors.

6.1. Major donors

We first consider the actual major donors with $d_1 = 0.1$. We compare the best informed LR to the LR obtained with various techniques: unrestricted integration assuming at most three contributors, restricted integration assuming three contributors with $0 \leq d_1 \leq 0.2$, $0.3 \leq d_2 \leq 0.5$, $0.6 \leq d_3 \leq 1$, unrestricted integration assuming at most four contributors, restricted integration assuming four contributors with $0 \leq d_1 \leq 0.2$, $0.4 \leq d_2 \leq 0.6$, $0.6 \leq d_3 \leq 1$, $0.8 \leq d_4 \leq 1$, and the point estimates $LR_{d_1,d_2,d_3}(M, g)$ for $(d_1, d_2, d_3) \in \{(0.1, 0.1, 0.1), (0.2, 0.2, 0.2), (0.3, 0.3, 0.3), (0.05, 0.5, 0.8), (0.2, 0.3, 0.8)\}$ where the last two choices represent reasonably close estimates of the dropout probabilities. For all these methods, we have compared Log_{10}(calculated LR)/Log_{10}(bestinformed LR). The results are presented in Fig. 8.

We conclude from Fig. 8, analogous to for the major donors in two-person mixtures (cf. Fig. 3) that the LR obtained by integration performs much better than evaluating the LR assuming specific probabilities of dropout that are equal for both contributors. We again obtain a weight of evidence that is closer to the best informed weight of evidence if we make no assumptions at all other than that the trace has at most four contributors, than if we evaluate the trace with the correct number of contributors, even if we use quite accurate estimates for the probability of dropout for all donors. We notice also that restricted integration performs slightly better than unrestricted integration but not to a great extent: as we already mentioned in Section 2.4, this is because the values that are discarded in the restricted integration, are among the least plausible given the mixture data and therefore their contribution to the LR obtained by unrestricted integration is relatively small.

6.2. Middle donors

We next consider the actual middle donors with $d_2 = 0.4$. We compare the best informed LR to the LR obtained with the same techniques as for the major donor except that we consider as point estimates the values $(d_1, d_2, d_3) \in \{(0.3, 0.3, 0.3), (0.4, 0.4, 0.4), (0.5, 0.5, 0.5), (0.05, 0.5, 0.8), (0.2, 0.3, 0.8)\}$, i.e., for the equal dropout choices we take a value around the actual value for this contributor. The results are presented in Fig. 9.

Similar to the situation for the two-person mixtures with $d_1 = d_2 = 0.3$, we see again that the method by which we assess the weight of evidence makes less difference than for the major donors. Integration, especially restricted three-person integration still has the best results in the sense that almost all weight of evidence is consistently conserved with the smallest variability across the simulated mixtures. We note that the probability of dropout of the contributor we look for now, is the average of that of all donors, and this may explain why we get relatively good results with LR's obtained from point estimates assuming equal dropout probabilities for all donors: the mixture likelihood will be optimal around the average probability of dropout, if all d_i must be equal. So this corresponds more or less to a maximum likelihood estimate

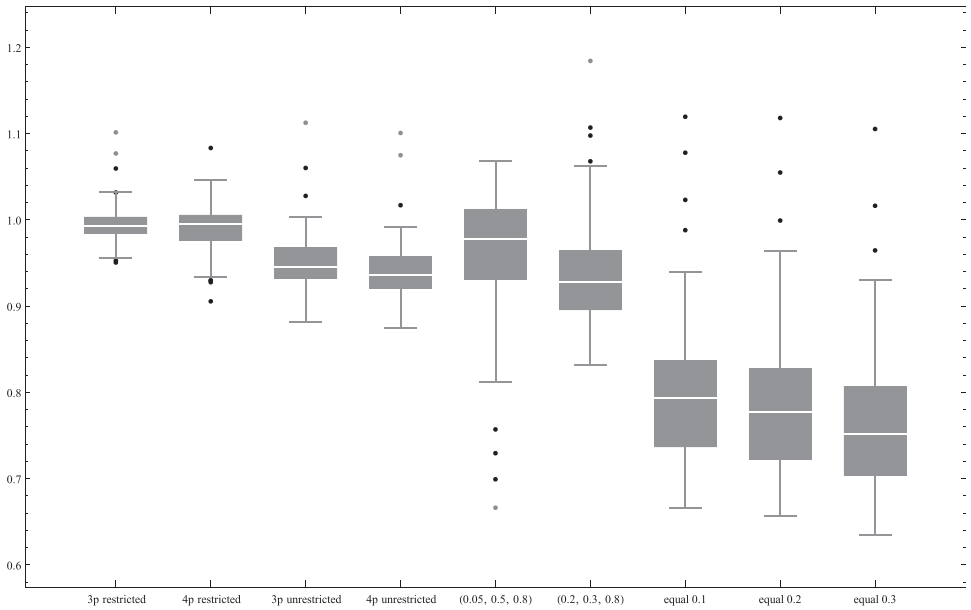


Fig. 8. Preserved fraction of weight of evidence for major donors in three-person mixtures with probabilities of dropout (0.1, 0.4, 0.7), when LR's are calculated using various methods.

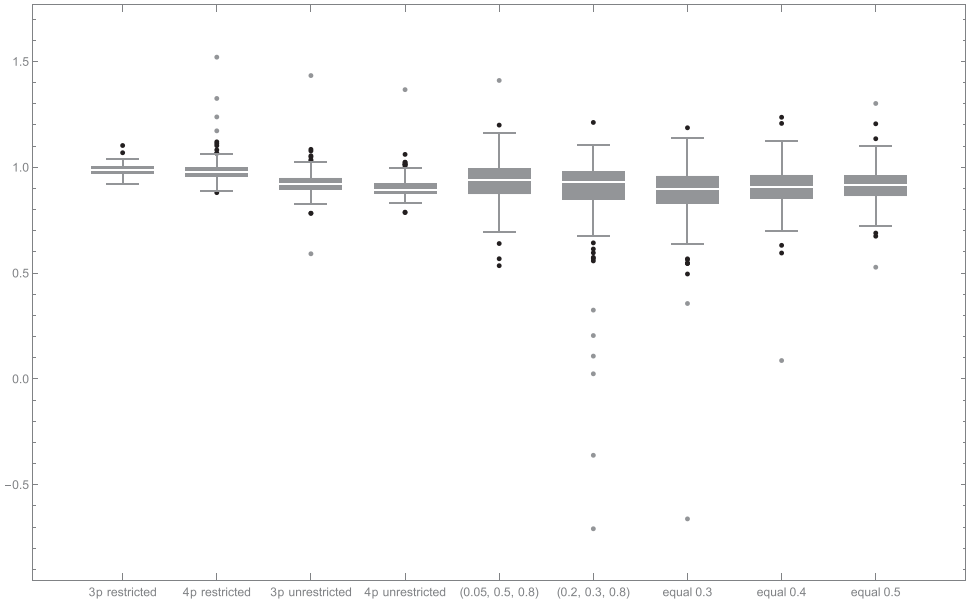


Fig. 9. Preserved fraction of weight of evidence for the middle donors in three-person mixtures with probabilities of dropout (0.1, 0.4, 0.7), when LR's are calculated using various methods.

among the equal dropout choices, and the maximum is achieved for the actual probability of dropout of the targeted donor.

6.3. Minor donors

We next consider the actual minor donors with $d_3 = 0.7$. We compare the best informed LR to the LR obtained with the same techniques as for the major donor, but again we use different point estimates for the LR's calculated with equal dropout probabilities for the three donors. The results are presented in Fig. 10.

We now see qualitatively the same behavior as for the minor donors in the two-person mixtures that we considered, cf. Fig. 4, but note the vertical scale: the results are now much more pronounced. All the equal dropout methods perform very poorly, and in many cases a weight of evidence indicating the opposite hypothesis as for the best informed weight of evidence is obtained (the results corresponding to negative scores). On the other hand, we also see that a good estimate of the probabilities of dropout leads to an accurate estimate of the weight of evidence. The integration methods perform more or less similarly to each other, although there are some differences. The restricted integration assuming three contributors preserves on average 98% of the weight of evidence. We provide some statistics in Tables 2 and 3.

Since it is generally a hard problem to assess the weight of evidence of a minor donor in a mixture and the weight of evidence for minor donors can be quite small, we also further compare these results by inspecting individual outcomes. In Fig. 11 we plot a bar chart, where for each of the 100 evaluated mixtures we plot a bar corresponding to the best informed weight of evidence (gray) and to the obtained estimate (black) using two of the methods just described: unrestricted integration with 3 contributors and the point estimate assuming dropout probabilities (0.7, 0.7, 0.7). We

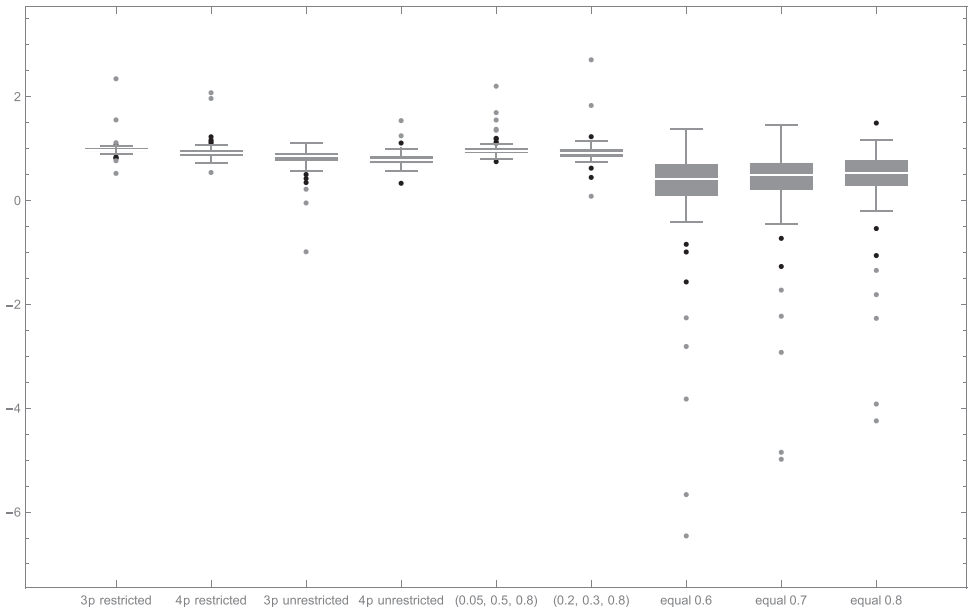


Fig. 10. Preserved fraction of weight of evidence for the minor donors in three-person mixtures with probabilities of dropout (0.1, 0.4, 0.7), when LR's are calculated using various methods.

Table 2
Statistics for preserved weight of evidence of minor donor with $d_3 = 0.7$ (based on 100 simulations) using integration method.

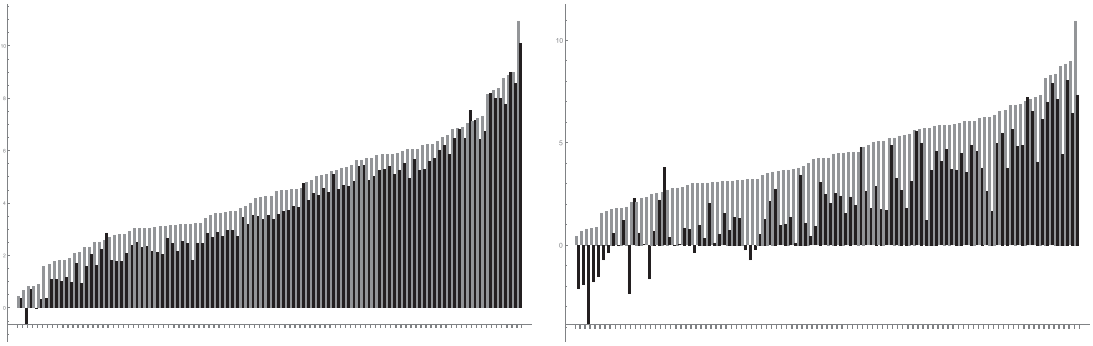
	3p restricted	4p restricted	3p unrestricted	4p unrestricted
Median	0.977	0.913	0.854	0.784
Mean	0.989	0.938	0.797	0.793
Standard deviation	0.1634	0.185	0.243	0.138
Proportion $\geq$150%	2/100	2/100	0	1/100
Proportion >50%	100/100	100/100	95/100	99/100
Proportion <0	0/100	0/100	2/100	0/100

Table 3
Statistics for preserved weight of evidence of minor donor with $d_3 = 0.7$ (based on 100 simulations) using point estimates.

	(0.05, 0.5, 0.8)	(0.2, 0.3, 0.8)	(0.6, 0.6, 0.6)	(0.7, 0.7, 0.7)	(0.8, 0.8, 0.8)
Median	0.958883	0.921783	0.422818	0.498781	0.534658
Mean	0.996103	0.944982	0.146827	0.274067	0.362815
Standard deviation	0.181105	0.245756	1.15468	0.966074	0.831308
Proportion $\geq$150%	3/100	2/100	0/100	0/100	1/100
Proportion >50%	100/100	98/100	43/100	50/100	54/100
Proportion <0	0/100	0/100	21/100	14/100	10/100

clearly see that the LR obtained by integration gives for almost all cases a slightly conservative estimate, whereas the weight of evidence is generally more strongly reduced, or even lost by the point estimate.

To conclude, as we have seen in this example, a good estimate of the probabilities of dropout allows for a good approximation of the best informed weight of evidence, but such an estimate will be hard to obtain from the mixture data since it involves a correct



(a) 3p unrestricted integration (b) Evaluated with $(0.7, 0.7, 0.7)$

Fig. 11. Obtained WoE by integration resp. by point estimate.

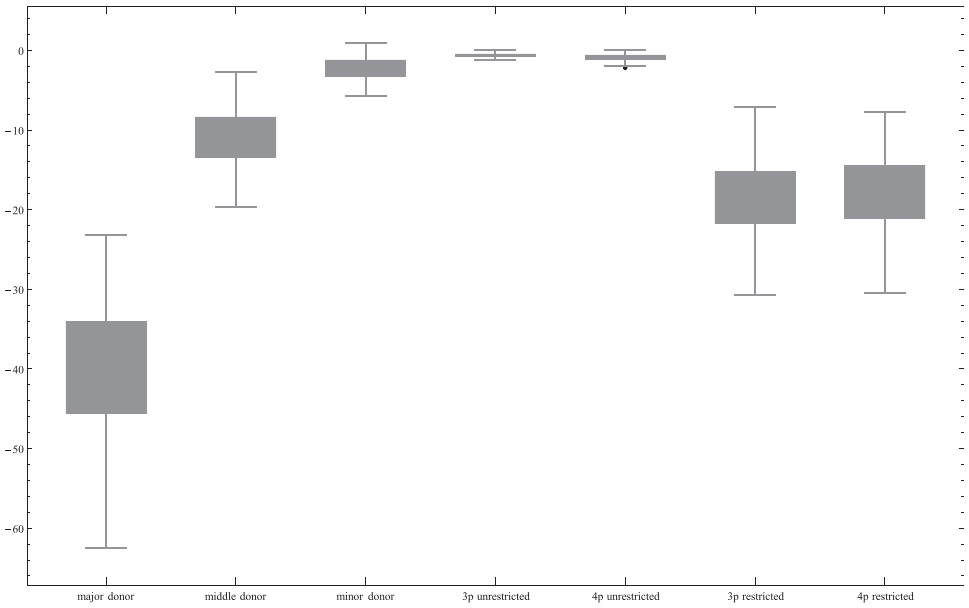


Fig. 12. Weight of evidence for non-contributors in three-person mixtures with probabilities of dropout (0.1, 0.4, 0.7), when LR's are calculated using various methods.

assignment of the number of contributors as well as a good estimate of the probability of dropout for each. If we, instead, estimate a range for each of the contributors and integrate the LR, we have an equally good and more feasible method. But even assuming the mixture to contain at most four contributors and not making any assumptions about their dropout probabilities still allows an accurate assessment of the weight of evidence.

6.4. Non-contributors

Finally we evaluate non-contributors. As was previously discussed, there is now no single best informed LR and we compare the obtained weights of evidence obtained with various methods: the three LR's obtained with probabilities of dropout equal to (0.1, 0.4, 0.7) for being the first, second or third donor; unrestricted integration with either three or four supposed contributors and restricted integration with three or four contributors where the dropout of the first donor, that we test for, is between 0 and 0.3 and the probabilities of dropout for the other donors are not restricted. Since, assuming three possible contributors, it is often impossible that a randomly drawn non-contributor can be a possible contributor with $c = 0$ since more than four alleles have been observed that the tested non-contributor does not have, we have taken $c = 1$ for the three-person evaluations. The resulting distributions of weight of evidence are displayed in Fig. 12.

If we first consider the best informed LR's we see that, as is logical, the exclusions are stronger if we test the non-contributor to be a more strongly present contributor.

When we consider the LR's obtained by integration, we see that if we do not restrict the probabilities of dropout, then we get LR's that are not very informative, also when considering three contributors. Contrary to the case where we considered non-contributors to two-person mixtures with $(d_1, d_2) = (0.1, 0.5)$ the least present donor here (with $d_3 = 0.7$) is less represented in the mixture, has two other donors to share alleles with, and we have used $c = 1$ which is relatively high. In that case, we again arrive at the statement that we cannot exclude that our non-contributor has contributed, but that it can only have been a very small amount. If we restrict the probability of dropout for the donor that we test for to be between 0 and 0.3 (and do not restrict the other contributors), then we get much stronger exclusions again, as can be seen from the boxplots corresponding to the restricted integration.

6.5. Determining the number of donors

To end the analyses of these mixtures, we again investigate as we did for the two-person mixtures (cf. Section 5.1.4), what LR's we obtain if we formulate hypotheses that investigate the number of contributors. Similarly to the analysis in Section 5.1.4, we compare the hypotheses of the mixture having three or four contributors both using unrestricted integration where no bounds on the dropout probabilities are imposed, and with an analysis where all contributors have at most 0.7 as probability of dropout. The resulting weights of evidence are presented in Fig. 13.

We see that, without restrictions on the probabilities of dropout, it is now even harder than for the two-person mixtures (cf. Fig. 6) to distinguish between the possibilities of three or four contributors: there is hardly any support either way. It is only when we define a contributor as someone who is represented to a minimum degree (here, a probability of dropout of at most 0.7), that we get support in favour of the mixture having three, rather than four, such contributors.

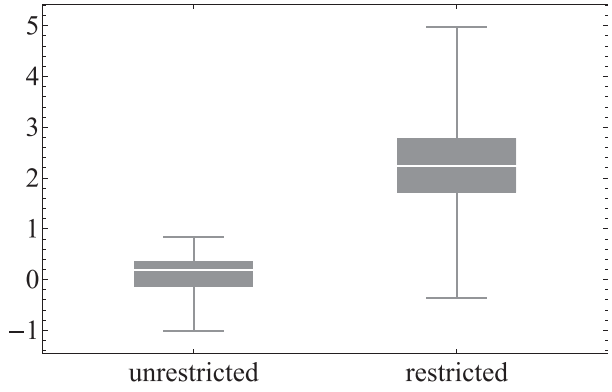


Fig. 13. Log_{10} (LR) for the hypotheses of the mixture having three versus four contributors.

7. Modeling issues

Throughout this article we have generated and processed all mixtures using (2.1). A possible point of criticism for this model is that it has as consequence that the probability D_i for a homozygous allele of contributor i to drop out is modeled to be equal to d_i^2, i.e., equal to the probability of two heterozygous alleles to drop out. In reality one would expect that D_i is strictly smaller than d_i^2 since (cf. [22]) conceivably each of the alleles on their own could yield a signal below the detection threshold whereas the cumulative signal exceeds it. Similarly it does not take into account the same phenomenon when alleles are shared between contributors. The main focus of this paper has been to show that integrating the LR recovers accurately the best informed LR, which we have demonstrated for the model with (2.1). Our results do not imply that (2.1) is an acceptable modeling assumption, but they do mean that if this model applies to the mixture data then we can accurately recover the best informed weight of evidence by integration.

Whether or not the model applies and to which extent is a separate problem which in full generality is beyond the scope and purpose of the paper, but in this subsection we nonetheless briefly digress in order to test the influence of modeling error. We will compare the LR obtained with (2.1) with the best informed LR when two modifications are made. First we consider an alternative for $D_i = d_i^2$ and as a second modification we consider degradation. Indeed, another consequence of (2.1) is that the probability of dropout d_i of contributor i is the same for all loci. In reality this probability could increase with increasing fragment length.

7.1. Homozygous dropout probability

First we look at an alternative for (2.1), defining

$$P_{\vec{d},c}(a \in \mathcal{M}|\vec{g}) = 1 - e^{-cp_a} \prod_{i:n_{i,a}=1} d_i \prod_{i:n_{i,a}=2} D_i. \tag{7.1}$$

Our definition (2.1) is obtained for the case $D_i = d_i^2$. For comparison we have taken

$$D_i = \frac{2^\beta d_i}{(1 + (2^\beta - 1)d_i)} \tag{7.2}$$

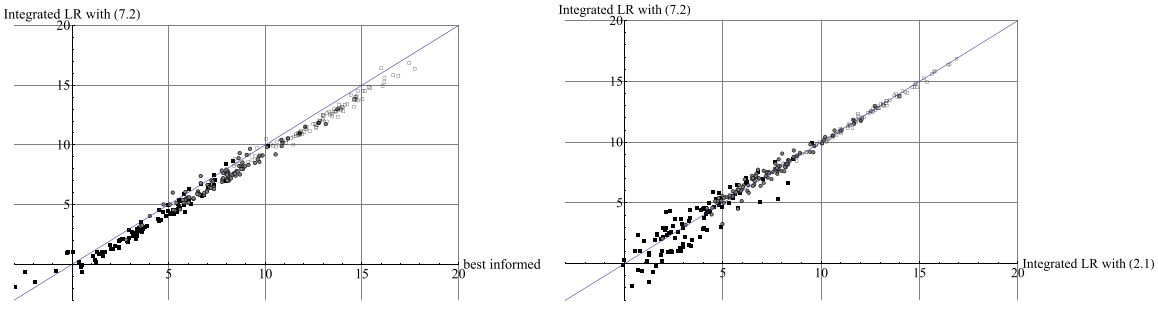
from [5] (see [13] for this presentation of the relation between D_i and d_i). Note that with this modification we have $D_i \to 0$ as $d_i \to 0$ and $D_i \to 1$ as $d_i \to 1$, which seems a basic requirement. For large probabilities of dropout, we have $D_i \ll d_i^2$. Thus, we expect that this model will have most impact compared to (2.1) for rather elevated dropout probabilities. To test this we have taken $\beta = -4.35$ as in [5] and sampled various mixture profiles according to (7.2) and then calculated the integrated LR both using D_i as in (7.2) and $D_i = d_i^2$ as in (2.1). We have done so for the donors of two-person mixtures with $(d_1, d_2) = (0.1, 0.5)$ and three-person mixtures with probabilities of dropout $(d_1, d_2, d_3) = (0.1, 0.4, 0.7)$.

We evaluate the obtained results in two ways: by comparing the integrated LR using (2.1) with the integrated LR using (7.2) and with the best informed LR (which is now also computed with (7.2) since the mixtures were generated using this formula). We plot the results in Fig. 14 which contains the results for the donors of the considered three-person mixtures. The results for the two-person mixtures are similar to those for the two most prominent donors of the three-person mixtures and are omitted.

We conclude that (2.1), while certainly not exactly correct, does not appear to be a harmful modeling assumption when major donors are evaluated, and that modeling care is needed when evaluating minor donors. We also see that, as was the case with (2.1), the integration method gives an accurate estimate of the best informed weight of evidence when (7.2) is used for the LR calculation. We also note however, that our choice to take $\beta = -4.35$ comes from [5]. This relation between D_i and d_i is based on logistic regression and not on a model, other data from other DNA multiplexes, PCR cycle numbers and machine settings may lead to other estimates (cf. [23]). We also note that the model of [13] for peak heights using gamma distributions, yields a smaller difference between D_i and d_i^2 than given by (7.2) if the mean of the observed peak heights is taken a proxy for the total amount of DNA (cf. [13]). If the model of [13] is accurate then this would mean that the simulation results in this subsection are obtained with an exaggeration of the difference between D_i and d_i^2.

7.2. Degradation

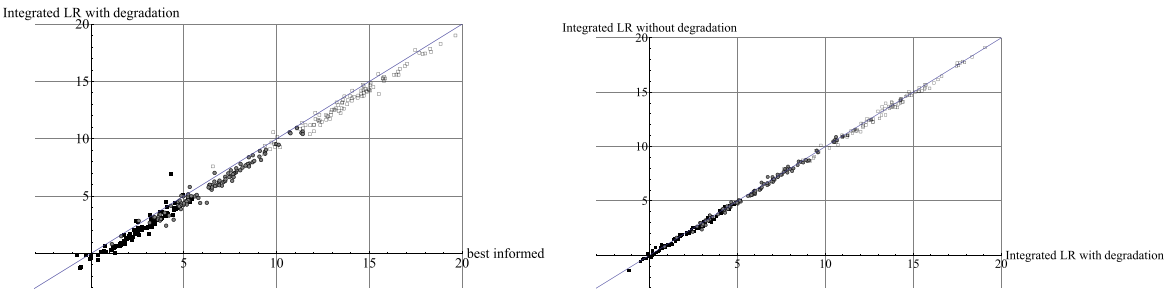
When dropout is a possibility, larger fragments may be more prone to dropout than shorter fragments. This leads to an increasing probability of dropout on each row of the obtained electropherogram. To model this, we have divided the NGM loci in four categories corresponding to their position on the dye lane where they are located. If a contributor has dropout probability d_i, we set a probability of dropout for a locus in position $p \in \{1, 2, 3, 4\}$ equal to $d_i^{(r^{p-1})}$ where $0 < r \leq 1$ is a parameter describing the degree of degradation. If $r = 1$ then we retrieve (2.1). For our simulations we have used $r = 0.8$. This means, for example, that if $d_i = 0.1$ then on the dropout probabilities on a dye lane increase from 0.1 for the locus in the first position to about 0.31 for the locus on the fourth position. A reason to choose this model is that if our degradation model is such that the heterozygous dropout probability on locus L can be written as $g(d_i, L) = d_i^{f(L)}$ for some function f of the loci, then this is compatible with (2.1) in the sense



(a) LR's obtained by integration using (7.2) and best informed LR's

(b) LR's obtained by integration using either (2.1) or (7.2)

Fig. 14. Obtained LR's for donors of mixtures that are generated with (7.2) with $\beta = -4.35$ and heterozygous probabilities of dropout (0.1, 0.4, 0.7). Empty squares correspond to the donor with $d_1 = 0.1$, gray circles to the donor with $d_2 = 0.4$ and filled black squares to the donors with $d_3 = 0.7$.



(a) LR's obtained by integration using (7.2) and best informed LR's

(b) LR's obtained by integration with and without degradation

Fig. 15. Obtained LR's for donors of mixtures that are generated with degradation and heterozygous probabilities of dropout (0.1, 0.4, 0.7). Empty squares correspond to the donor with $d_1 = 0.1$, gray circles to the donor with $d_2 = 0.2$ and filled black squares to the donors with $d_3 = 0.7$.

that $g(d_i, L)^2 = (d_i^{f(L)})^2 = (d_i^2)^{f(L)} = g(d_i^2, L)$. Again we have generated mixtures according to this model, computed the best informed LR according to the generating model, and computed the integrated LR with this model as well as with (2.1). The results are shown in Fig. 15.

Again, the results for the two-person mixtures are similar and we omit them. We see from Fig. 15 that the integrated LR is virtually the same whether or not degradation is taken into account in the evaluation or not, for these mixtures. Of course in reality it would be hard to estimate the degradation parameter, and one could argue how to model this phenomenon better, but for these cases this would not have been necessary as the LR obtained by integration is unaffected by it. If it can be satisfactorily modeled as a refinement of the semi-continuous model described here, it would be preferable to implement this, but from these results it seems that the lack of degradation modeling does not appear to seriously affect the capacity to accurately recover the best informed weight of evidence.

8. Discussion

For weight of evidence calculations on DNA mixtures involving a probabilistic model, it may be hard to estimate the parameters that are mixture-specific, such as the number of contributors, their relative contribution, etc. For the semi-continuous model discussed here, the calculated likelihood ratio will depend on the chosen probabilities of dropout for each donor and these are hard to estimate precisely. However, the estimation of these parameters is much aided by a likelihood-driven approach taking the allele frequencies of the observed alleles into account, as well as the reproducibility of the alleles across several independent replicates. Especially taking the allele frequencies into account is hard for a human interpreter, but straightforward for a computer. In this paper, we have investigated how well we can recover the LR that would have been obtained if all relevant mixture parameters would be known, if we instead start with a uniform prior on the dropout probabilities and determine, based on the mixture data, the (posterior to the mixture data) probability density function for these parameters. We then integrate the LR with respect to this probability density function. This way, a LR that belongs to more plausible parameters puts in more weight into the final result than a LR obtained for unrealistic values. It is possible, but not necessary, for a human interpreter to assess the number of contributors and restrict their probabilities of dropout to some set of values that are deemed the most likely. If these restrictions are accurate, then they will not influence the resulting LR to a large extent, since only implausible values are discarded that do not contribute much.

We compare this method with the calculation of point estimates $LR_{\vec{d},c}(M, g)$ where the LR is calculated within the probabilistic model with parameters $\vec{d}, c$. This is probably the most common application of the semi-continuous model in casework among forensic laboratories, and then especially the models where all unknown donors must have the same probability of dropout, since some semi-continuous model implementations do not support other choices.

We have seen that the integration method, although it involves little or no human interpretation, gives an accurate estimate of the weight of evidence for true donors of the mixture. Regardless of whether the tested person is an actual major or minor donor, the obtained LR by integration is very close to the best informed LR even if the allowed number of contributors exceeds the actual number. Whether or not the point estimates perform better or not, depends of course very much on the quality of these estimates. In our examples, even point estimates that are close do not outperform the integrated LR.

Calculations assuming equal dropout probabilities for all unknown donors can be a cause of a large loss of information, without getting rid of the risk of occasional overestimation of the weight of evidence. In these examples the equal dropout approach only gives accurate results when the actual dropout probabilities are such that their average is equal to the actual probability of dropout for the donor that we consider, and not when the donor that we consider has a higher or lower than average dropout probability. Integration over the probability of dropout, assuming that all contributors have the same, does not solve this problem as was already made clear in the example in Section 3.

For non-contributors, the results of the integration are also perfectly reasonable. If the allowed number of contributors is equal to the actual number, then – depending on the dropout probabilities of the real donors – it may be clear that there is a counter-indication for the PoI to be any of the contributors and we will obtain a LR corresponding mostly to the exclusion of the donor that is hardest to exclude (the one with the highest probability of dropout). It may be the case that the donor with the highest dropout probability is hard to recognize, and then the LR's for non-contributors become less informative, as we saw for the three person mixtures. If we allow more contributors than there actually are, then the LR is typically almost uninformative and only slightly smaller than one. This corresponds to the conclusion that, although there is no indication that the PoI has contributed, it still cannot be ruled out that he has contributed such a tiny amount of DNA to the trace that it cannot be found back in the trace. Such a statement only says that no indication has been found that the PoI has contributed to the trace profile, even if only minute contributions are considered, but that of course such a contribution cannot be

ruled out either. This is a very reasonable conclusion, if indeed we are making a statement about a possible non-zero contribution. We can make a stronger statement by restricting the domain of dropout variables to only contain values that are not too high, such as the interval [0, 0.5]. Then the LR for a non-contributor strongly points towards non-contribution. This then means that for this PoI, we have strong evidence against a significant contribution. This way, the method presented here naturally leads us to consider when we actually consider someone to be a contributor.

The method thus has several desirable properties. It is a method that resembles human interpretation in the sense that it holds everything to be possible a priori, but the end result is most influenced by what appears to be the best explanation for the mixture data (where the quality of the explanation, i.e., the mixture likelihood, is obtained without comparing it to the PoI). Second, it is more effective than human interpretation since allele frequencies are taken into account and the integration ranges over all possible parameter values. It is also an objective method in the sense that no estimate needs necessarily to be made regarding the number of contributors (other than a maximum) nor about their respective contributions. Of course, it is possible to use estimates for the dropout probabilities, but ranges rather than point estimates suffice.

We can also view these results as belonging to a validation of the semi-continuous method, since it shows that it is possible to accurately retrieve the weight of evidence that it is carried by the mixture: we have shown that we can closely recover the best informed LR by integration, where we only make use of the mixture data. Moreover we have shown that, at least for major donors, the LR obtained by integration using (2.1) still performs well as an estimator of the best informed weight of evidence, even if the mixture data have not been obtained in agreement with the model, as was described in Section 7. Of course, it would be preferable to make the model as realistic as possible provided we know how to do so. On the other hand, such improvements can bring the additional complication of being specific to the DNA typing technology (multiplex, cycle number, CE settings, etc.) and trace type (degradation, inhibition, template amount, etc.). This brings us to the question of how the results in this paper relate to those obtained with a continuous model. Without being an expert on these systems, we believe that it must be hard in general to build a model predicting peak height distributions applicable to all profiles in view of the variation that we just mentioned. The semi-continuous model used in this paper discards the peak height information and will be less powerful than a continuous model if the peak height model it uses is accurate. On the other hand, the information that the semi-continuous model with integration does process, is processed quite correctly as long as it can be assumed that there exists a probability of dropout that is contributor-specific and does not vary among replicates. The relation between semi-continuous models and continuous models is analogous to the relation between inclusion probabilities and likelihood ratios for mixtures without artefacts obtained by a binary model (i.e., the semi-continuous model without dropout or drop-in). The inclusion probability uses less information (being calculated from the mixture and not using the profile of the person of interest), but (their inverse) can be regarded as the weight of evidence when the evidence is considered to be only the fact that there is a matching person of interest without disclosing that person's genotype (cf. [16]). A LR calculation with the binary model uses more information and therefore provides a better informed weight of evidence which is, for true donors, in expectation more than obtained from exclusion probabilities. Similarly, the ideal continuous model will outperform the semi-continuous model, but the advantage of the semi-continuous model is its simplicity which makes it widely applicable across trace types and typing technologies, as a reliable and unbiased estimator of the weight of evidence that is obtained from peak presence or absence. Therefore the results obtained with method of this paper could potentially serve in a comparison of results with continuous models to validate the latter (similar to approaches in [24,14]).

As we have seen, restricting the integration domain to a subset of most plausible values is, for actual donors, of some use, but mostly in the sense of providing some computational advantage when numerical integration is carried out as we have done here. The difference in obtained LR is small since discarding the least plausible values corresponds to little change in obtained information. However, it can be useful to not only obtain evidence in favor of contribution, but to also obtain evidence in favour of contribution within certain ranges of the probability of dropout. That way a statement can be made both on the weight of evidence and on the amount of representation of the person of interest's profile in the mixture.

We have also investigated whether we can use this approach to estimate the number of contributors to a mixture. However, when we ask this question, we again come across the need to first decide when we actually consider a person to be a contributor. In DNA mixture profiles exhibiting dropout, the whole concept of number of contributors becomes less meaningful than it is for mixture without dropout. We have indeed seen that, if we do not make any restrictions on the probabilities of dropout, then the assessment of the number of contributors becomes much harder than if we set a maximum value on the probability of dropout for each.

Finally, let us comment on some aspects that we have not considered here. A first omission is that we have not treated possible kinship between the person of interest and the actual donor that we are targeting in the LR calculation. However, of course also likelihood ratios taking such a possibility into account can in principle be subjected to the same method.

In this manuscript we have not used a θ-correction. The reason that we have not applied it is that we have not sampled the profiles of the contributors with a θ-correction, and so the best informed LR is the one with $\theta = 0$. The purpose of this manuscript is to show that by integrating the LR, we obtain a weight of evidence that is very close to the weight of evidence that we would obtained had we known all relevant parameters. Of course, it is possible to apply a θ-correction, obtain the LR algebraically again, and integrate; we have in fact implemented this in the Mathematica script. A θ-correction has a conservative effect on the likelihood ratio, but in many cases this is also true for taking estimates of the LR compared to the best informed LR. We did not want to confound these two effects in this research. Depending on the chosen value for θ the result will be conservative within the subpopulations of the population from which the allele frequency database is taken or even in other populations (cf. [25,26]).

References

[1] H. Haned, K. Slooten, P. Gill, Exploratory data analysis for the interpretation of low template DNA mixtures, Forensic Sci. Int.: Genet. 6 (6) (2012) 762–774.

[2] K. Inman, N. Rudin, K. Cheng, C. Robinson, A. Kirschner, L. Inman-Semerau, K. Lohmueller, Lab retriever: a software tool for calculating likelihood ratios incorporating a probability of drop-out for forensic DNA profiles, BMC Bioinf. 16 (298) (2015).

[3] D. Balding, Evaluation of mixed-source, low-template DNA profiles in forensic science, Proc. Natl. Acad. Sci. U. S. A. 110 (30) (2013) 12241–12246.

[4] K. Slooten, Distinguishing between donors and their relatives in complex DNA mixtures with binary models, Forensic Sci. Int.: Genet. 21 (2015) 95–109.

[5] T. Tvedebrink, P. Eriksen, H. Mogensen, N. Morling, Estimating the probability of allelic drop-out of STR alleles in forensic genetics, Forensic Sci. Int.: Genet. 3 (2009) 222–226.

[6] T. Tvedebrink, P.S. Eriksen, H.S. Mogensen, N. Morling, Statistical model for degraded DNA samples and adjusted probabilities for allelic drop-out, Forensic Sci. Int.: Genet. 6 (2012) 97–101.

[7] J. Buckleton, H. Kelly, J.-A. Bright, D. Taylor, T. Tvedebrink, J. Curran, Utilising allelic dropout probabilities estimated by logistic regression in casework, Forensic Sci. Int.: Genet. 9 (2014) 9–11.

[8] J. Buckleton, J. Curran, P. Gill, Towards understanding the effect of uncertainty in the number of contributors to DNA stains, Forensic Sci. Int.: Genet. 1 (1) (2007) 20–28.
[9] H. Haned, L. Pène, J. Lobry, A. Dufour, D. Pontier, Estimating the number of contributors to forensic DNA mixtures: does maximum likelihood perform better than maximum allele count? J. Forensic Sci. 56 (1) (2011) 23–28.
[10] J. Perez, A. Mitchell, N. Ducasse, J. Tamariz, T. Caragine, Estimating the number of contributors to two-, three-, and four-person mixtures containing DNA in high template and low template amounts, Croat. Med. J. 52 (3) (2011) 314–326.
[11] A. Biedermann, S. Bozza, K. Konis, F. Taroni, Inference about the number of contributors to a DNA mixture: comparative analyses of a Bayesian network approach and the maximum allele count method, Forensic Sci. Int.: Genet. 6 (6) (2012) 696–698.
[12] D. Taylor, J. Bright, J. Buckleton, The interpretation of single source and mixed DNA profiles, Forensic Sci. Int.: Genet. 7 (5) (2013) 516–528.
[13] R. Cowell, T. Graversen, S. Lauritzen, J. Mortera, Analysis of forensic DNA mixtures with artefacts, J. R. Stat. Soc. Ser. C 64 (1) (2014) 1–48 (with discussion).
[14] C. Steele, M. Greenhalgh, D. Balding, Evaluation of low-template DNA profiles using peak heights, Stat. Appl. Genet. Mol. Biol. 15 (5) (2016) 431–445.
[15] K. Slooten, Familial searching on DNA mixtures with dropout, Forensic Sci. Int.: Genet. 22 (2016) 128–138.
[16] K. Slooten, T. Egeland, Exclusion probabilities and likelihood ratios with applications to mixtures, Int. J. Leg. Med. 130 (1) (2016) 39–57.
[17] D. Balding, R. Nichols, DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands, Forensic Sci. Int. 64 (1994) 125–140.
[18] A. Westen, et al., Comparing six commercial autosomal STR kits in a large Dutch population sample, Forensic Sci. Int.: Genet. 10 (2014) 23–55.
[19] I. Good, Studies in the history of probability and statistics. XXXVII A.M. Turing's statistical work in World War II, Biometrika 66 (2) (1979) 393–396.
[20] F. Taroni, S. Bozza, A. Biedermann, C. Aitken, Dismissal of the illusion of uncertainty in the assessment of a likelihood ratio, Law Probab. Risk 15 (1) (2016) 1–16.
[21] M.J. Sjerps, I. Alberink, A. Bolck, R.D. Stoel, P. Vergeer, J.H. van Zanten, Uncertainty and LR: to integrate or not to integrate, that's the question, Law Probab. Risk 15 (1) (2016) 23–29.
[22] D. Balding, J. Buckleton, Interpreting low template DNA profiles, Forensic Sci. Int.: Genet. (4) (2009) 1–10.
[23] T. Tvedebrink, P.S. Eriksen, M. Asplund, H.S. Mogensen, N. Morling, Allelic drop-out probabilities estimated by logistic regression-further considerations and practical implementation, Forensic Sci. Int.: Genet. 6 (2012) 263–267.
[24] J.-A. Bright, I. Evett, D. Taylor, J. Curran, J. Buckleton, A series of recommended tests when validating probabilistic DNA profile interpretation software, Forensic Sci. Int.: Genet. 14 (2015) 125–131.
[25] C. Steele, D. Balding, Choice of population database for forensic DNA profile analysis, Sci. Justice 54 (6) (2014) 487–493.
[26] J. Buckleton, J. Curran, J. Goudet, D. Taylor, A. Thiery, B. Weir, Population-specific F_{ST}-values for forensic STR markers: a worldwide survey, Forensic Sci. Int.: Genet. 23 (2016) 91–100.