

## Two or three contributors of DNA mixtures? – Some practical considerations

Amke Caliebe

Institute of Medical Informatics and Statistics, Kiel University, University Hospital Schleswig-Holstein, Brunswiker Strasse 10, 24105 Kiel, Germany

### ARTICLE INFO

#### Keywords:

Number of contributors  
Bayesian approach  
DNA mixture  
Prior distribution  
Nuisance parameter  
Likelihood ratio

### ABSTRACT

The number of contributors is hard to determine in DNA mixture profiles. Here, we deal with the special but frequent case that either two or three contributors are possible. In fact, it might happen that two contributors can explain the number of alleles seen but that three contributors are necessary if a specific person of interest is to be included in the mixture. Then the likelihood ratio assuming two contributors will be zero while the likelihood ratio for three contributors may be large. We evaluate this situation and offer suggestions on how to arrive at an overall likelihood ratio. To exemplify our line of reasoning we use an example proposed by Biedermann, Taroni and Thompson.

### 1. Introduction

In DNA mixtures it is often not clear which number of contributors (NoC) to choose. There has been a debate ongoing on whether the NoC should be the same or not for the two hypotheses to be compared in a likelihood ratio (LR) calculation [2–5]. This subject has been treated thoroughly in [6] where it was clarified under which conditions the same or different NoC should be selected. In this study we will focus on a situation which arises frequently in casework for mixed DNA profiles: The trace profile shows at most four alleles for each locus. Then in a common (though overly simplified) approach, two contributors might be assumed. Later, the profile of a person of interest (PoI) is derived and it turns out that under the condition that this PoI is a contributor, three contributors are necessary to explain the mixed trace. Is it then reasonable to calculate the LR for three contributors (as the prosecution might favour) or have two contributors to be chosen (as the defense might insist)? This problem will be considered in the following and practical advice about how to handle this scenario will be given. In this manuscript, we will frequently refer to a highly illustrative example given by Biedermann, Taroni and Thompson [1] that will be stated now.

#### 1.1. Hat Example of Biedermann, Taroni and Thompson

The so-called ‘Hat Example’ was described in [1] in the following way: “Two brothers B and C were prosecuted for murder. According to the prosecution theory, the brothers entered a store where B wrestled with the clerk and C shot the clerk with a handgun. A video surveillance tape showed the crime occurring. Although the faces of the assailants could not be seen, the video revealed that the shooter had worn a hat

that was found at the crime scene. Key evidence in the case was the DNA profile for a mixed stain found on the hat [...]” [1] This seven-loci stain had three or four alleles to each locus, so a NoC of two might be acceptable. When comparing suspect C to the stain, however, it was found that C was homozygous for allele 13 for locus D5S818 whereas the trait consisted of alleles {8, 11, 12, 13}. If we denote by  $H_p$  the hypothesis that C contributed to the trace and  $D$  are the profile data, then three contributors must be assumed under  $H_p$  for a non-zero likelihood  $L(D) = P(D|H_p)$  (under a simple model without using peak heights and no drop-in or drop-out). A similar problem was noted for locus D21S11.

### 2. Methods

In [6] the NoC was treated as a nuisance parameter in Bayesian fashion. Let us suppose that we have a minimal NoC  $n_{min}$  and a maximal NoC  $n_{max}$  and that all NoCs between  $n_{min}$  and  $n_{max}$  are possible under both  $H_p$  and  $H_d$ . Then the overall LR  $LR(D) = P(D|H_p)/P(D|H_d)$  can be written as a weighted average over the  $LR^{(n)}$ s for  $n$  contributors:

$$LR(D) = \sum_{n=n_{min}}^{n_{max}} LR^{(n)}(D)P(\text{NoC} = n|D, H_d) \frac{P(\text{NoC} = n|H_p)}{P(\text{NoC} = n|H_d)}. \quad (1)$$

Here,  $P(\text{NoC} = n|D, H_d)$  is the a posteriori probability to have  $n$  contributors when the PoI is not contributing and given the data  $D$ . The expressions  $P(\text{NoC} = n|H_p)$  and  $P(\text{NoC} = n|H_d)$  refer to the prior probabilities to have  $n$  contributors either under  $H_p$  or  $H_d$ . For these prior probabilities the genetic information  $D$  is not taken into account and they are solely determined by other case circumstances. Although they might be different for  $H_p$  and  $H_d$ , this is only in some cases sensible [6]

E-mail address: [caliebe@medinfo.uni-kiel.de](mailto:caliebe@medinfo.uni-kiel.de).

<https://doi.org/10.1016/j.fsigss.2019.09.090>

Received 17 September 2019; Accepted 26 September 2019

Available online 16 October 2019

1875-1768/ © 2019 Elsevier B.V. All rights reserved.

and we will restrict ourselves to the case that  $P(\text{NoC} = n|H_p) = P(\text{NoC} = n|H_d)$ . Then the prior probabilities in Eq. (1) cancel out and we obtain the simplified equation

$$LR(D) = \sum_{n=n_{\min}}^{n_{\max}} LR^{(n)}(D)P(\text{NoC} = n|D, H_d).$$

Let us now return to the case of either two or three contributors. For simplicity we assume that other NoCs are not possible. Then the sum consists only of two terms

$$LR(D) = LR^{(2)}P(\text{NoC} = 2|D, H_d) + LR^{(3)}P(\text{NoC} = 3|D, H_d).$$

We are interested in the situation that under  $H_p$  only two contributors are possible. That means that  $LR^{(2)} = 0$  and we obtain the central equation

$$LR(D) = LR^{(3)}P(\text{NoC} = 3|D, H_d) \tag{2}$$

which we will investigate further in order to get leads about how to deal with the ambiguity in the NoC.

### 3. Results

When handling the two versus three contributors situation, we have two LRs to consider:  $LR^{(2)} = 0$  and  $LR^{(3)}$ . The latter might be large. Which LR should be reported? When looking at the overall LR of Eq. (2), we see that  $LR^{(3)}$  is weighted by the probability of having three contributors derived from the trace data when the PoI has not contributed  $P(\text{NoC} = 3|D, H_d)$ . This probability is between 0 and 1, thus LR is between 0 and  $LR^{(3)}$ . For evaluating whether LR is nearer to 0 or to  $LR^{(3)}$ , the weighting factor  $P(\text{NoC} = 3|D, H_d)$  has to be (at least roughly) estimated. To do that, we write this probability in a different kind of way:

$$P(\text{NoC} = 3|D, H_d) = \frac{P(\text{NoC} = 3)LR^d}{P(\text{NoC} = 3)(LR^d - 1) + 1}$$

where  $LR^d = P(D|\text{NoC} = 3, H_d)/P(D|\text{NoC} = 2, H_d)$ . We see that  $P(\text{NoC} = 3|D, H_d)$  depends on the two terms  $LR^d$  and  $P(\text{NoC} = 3)$  and we will consider these separately in the following.

#### 3.1. The likelihood ratio $LR^d$

$LR^d = P(D|\text{NoC} = 3, H_d)/P(D|\text{NoC} = 2, H_d)$  compares the probability of the trace data under either two or three contributors in the situation where the PoI is not part of the mixture.  $LR^d$  will be large if three contributors are much more likely than two when looking at the mixture. Therefore, it is not surprising that for large values of  $LR^d$ , also the weighting factor  $P(\text{NoC} = 3|D, H_d)$  gets large, i.e. near to one and the overall LR is near to  $LR^{(3)}$ .  $LR^d$  is a LR for two different propositions, namely two or three contributors and can therefore be readily calculated from the data.

#### 3.2. The prior probability $P(\text{NoC} = 3)$

Whereas for  $LR^d$  the mixture profile D is essential,  $P(\text{NoC} = 3)$  is the prior probability of having three contributors to the trace without using the genetic information. The larger the prior probability for three contributors is, the nearer the overall LR will be to  $LR^{(3)}$ .  $P(\text{NoC} = 3)$  can only be derived from the non-genetic evidence and case circumstances. As such, it cannot be calculated and can at best be loosely estimated in contrast to  $LR^d$ . Depending on the information available for a specific case, it might be possible to make inference about the order of  $P(\text{NoC} = 3)$ , e.g. by eye witness evidence or when it is known that only two persons have access to a special item (such as victim and PoI). In other circumstances, no information about  $P(\text{NoC} = 3)$  might be available. Then a prior of 1/2 could be one

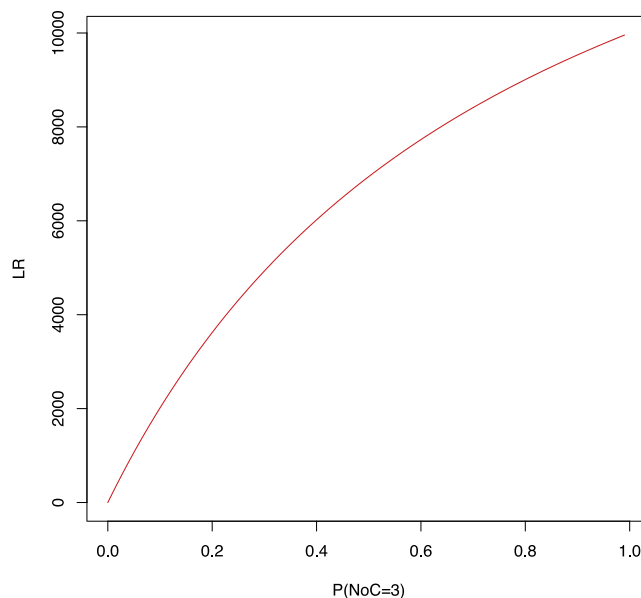


Fig. 1. LR as a function of  $P(\text{NoC} = 3)$  for the Hat Example, where  $LR^{(2)} = 0$ ,  $LR^{(3)} \approx 10,000$ ,  $LR^d = 2.27$ .

possibility. For sensitivity analysis, a number of sensible priors should be taken into account.

#### 3.3. Returning to the Hat Example

Let us now return to the Hat Example of Biedermann, Taroni and Thompson and try to apply our previous considerations. For this mixture stain, we have  $LR^{(2)} = 0$  and  $LR^{(3)} \approx 10,000$  [1]. The overall LR therefore is between 0 and 10,000. To assess the weighting factor  $P(\text{NoC} = 3|D, H_d)$ , we have to consider  $LR^d$  and  $P(\text{NoC} = 3)$ . As noted before,  $LR^d$  can be calculated from the data and is 2.27 in this case [1]. As usual, the determination of the prior probability  $P(\text{NoC} = 3)$  is much more difficult. Because we have a stain on a hat which can be touched by many or few people, no real prior information is available. Both two or three contributors are possible. The dependence of LR on the prior  $P(\text{NoC} = 3)$  is shown in Fig. 1. If we take  $P(\text{NoC} = 3) = 1/2$ , then a LR around 7000 will result. Because the prior distribution is uncertain and to be conservative, it would make sense to arrive at an overall LR which is somewhat lower, e.g. around 5000.

### 4. Discussion

This manuscript deals with a special case in the NoC discussion to make the results clear and accessible also to non-statisticians. This is also the reason why we apply the results to the elucidating Hat Example of Biedermann, Taroni and Thompson [1]. Of course, generally other NoCs are possible. The Hat Example regards a simple discrete model without drop-out or drop-in. If fully continuous models are considered (so-called ‘probabilistic genotyping’), the situation somewhat changes because, if too many contributors are chosen, then their contribution might be modelled as nearly zero. Another important simplification in this manuscript is that only one prior distribution for the NoC is applied, the same for  $H_p$  and  $H_d$ . Although this is a reasonable approximation for many cases, there are also situations where this is not appropriate. For more information on these subjects see [1,6].

### 5. Conclusion

Because drop-out, drop-in and minor contributions are possible for DNA mixtures, the NoC can never be determined without uncertainty.

Therefore, calculation of the LR for several NoCs is required. The probabilities of the data for different NoCs under  $H_d$  and the prior distributions of the NoC then influence the overall LR.

#### Conflict of interest statement

The author of this manuscript declares no conflict of interest.

#### Acknowledgements

The author would like to thank Klaas Slooten for many happy and fruitful discussions.

#### References

- [1] A. Biedermann, F. Taroni, W.C. Thompson, Using graphical probability analysis (Bayes Nets) to evaluate a conditional DNA inclusion, *Law Probab. Risk* 10 (2011) 89–121.
- [2] C.H. Brenner, Fairness in evaluating DNA mixtures, *Forensic Sci. Int. Genet.* 27 (2017) 186.
- [3] I.W. Evett, Another response to “About the number of contributors to a forensic sample”, *Forensic Sci. Int. Genet.* 28 (2017) e11.
- [4] P. Gill, A response to “About the number of contributors to a forensic sample”, *Forensic Sci. Int. Genet.* 26 (2017) e9–e13.
- [5] S. Presciuttini, T. Egeland, About the number of contributors to a forensic sample, *Forensic Sci. Int. Genet.* 25 (2016) e18–e19.
- [6] K. Slooten, A. Caliebe, Contributors are a nuisance (parameter) for DNA mixture evidence evaluation, *Forensic Sci. Int. Genet.* 37 (2018) 116–125.