Law, Probability and Risk (2016) **15**, 23–29
Advance Access publication on September 11, 2015 doi:10.1093/lpr/mgv005

Uncertainty and LR: to integrate or not to integrate, that's the question

M. J. Sjerps†
Netherlands Forensic Institute, P.O. Box 24044, 2490 AA The Hague, the Netherlands; Korteweg-de Vries Institute for Mathematics, University of Amsterdam, P.O. Box 94248, 1090 GE Amsterdam, The Netherlands

I. Alberink
Netherlands Forensic Institute, P.O. Box 24044, 2490 AA The Hague, the Netherlands

A. Bolck
Netherlands Forensic Institute, P.O. Box 24044, 2490 AA The Hague, the Netherlands

R. D. Stoel
Netherlands Forensic Institute, P.O. Box 24044, 2490 AA The Hague, the Netherlands

P. Vergeer
Netherlands Forensic Institute, P.O. Box 24044, 2490 AA The Hague, the Netherlands

and

J. H. van Zanten
Korteweg-de Vries Institute for Mathematics, University of Amsterdam, P.O. Box 94248, 1090 GE Amsterdam, The Netherlands

[Received on 22 May 2015; accepted on 22 May 2015]

Taroni *et al.* (2016) discuss the controversial issue of parameter uncertainty in the context of forensic evidence evaluation. Although we share with the authors the main idea that the likelihood ratio (LR) framework is the best method for evaluating forensic evidence, we have a different view on this issue. The core question is: does it make sense to consider the uncertainty attached to a calculated value of the LR, and consequently, should we report a single value for the LR or in addition address its uncertainty? Taroni *et al.* (2016) argue for reporting a single value based on a 'full-Bayesian' approach, and accuse anyone who considers the uncertainty of an LR of 'misconception of basic principles' and 'abuse of language'. However, their arguments presented as facts or logic are in fact choices or opinions. Furthermore, reporting a single number for the LR deprives the legal justice system of essential information needed to assess the reliability of the evidence. Therefore, we argue that forensic scientists should not only report an LR value, but also address its uncertainty and we explain why this is not a misconception or abuse of language.

Keywords: Likelihood ratio; uncertainty; forensic science; Bayes factor; estimation; reliability; evidential value.

†Corresponding author. Email: m.sjerps@nfi.minvenj.nl

1. Introduction

In their paper, 'Dismissal of the illusion of uncertainty in the assessment of a likelihood ratio', Taroni *et al.* (2016, abbreviated here as TBBA) give their view on a controversial issue: dealing with parameter uncertainty when evaluating evidence in a forensic context. This issue is important because it is at the basis of a popular methodology for interpreting and evaluating forensic evidence, and has consequences for the way that this evidence is reported to the legal justice system. It has been debated for a long time at conferences such as the ICFIS series (International Conference on Forensic Inference and Statistics). TBBA explain the issue and their view in great detail, and we appreciate the thoroughness of their argumentation.

We agree with TBBA for an important part, but disagree with them on a crucial issue, and consequently have a different view on how forensic evidence should be reported. In this reaction on TBBA, we start with identifying our common ground. Subsequently, we identify where our opinions diverge, and argue that some of the arguments presented in TBBA as facts or logical consequences are in fact choices or opinions. We end by explaining why we prefer other choices for reporting forensic evidence.

2. Common ground

We would like to emphasize that TBBA discuss a topic that we consider as a small detail in the larger context of assessing evidential strength in forensic science. We share with TBBA the most important ideas in this area, some of which have been developed by TBBA, and we generally appreciate their previous work. We list some of these ideas below just to clarify the discussion in the next section:

- The LR framework is the optimal framework for assessing the strength of forensic evidence.
- For probabilistic reasoning in the forensic domain, the subjective definition of the concept 'probability' is often useful. This includes the idea that such probabilities are personal and conditional on the available information and premises.
- Reporting a lower bound or interval for a LR requires a rather arbitrary choice of a limit or level, e.g. 95% or 99% credible or confidence limit.
- Performance measures of a LR method are useful for assessing the general validity of an analytical methodology with an associated LR. In a specific case, the strength of the particular observations is evaluated in terms of a likelihood ratio.

3. Different views

3.1 *An important choice: dealing with parameter uncertainty*

Despite this large area of common ground, we have quite a different view on the central issues raised in TBBA: does it make sense to consider the uncertainty attached to a calculated value of the LR, and consequently, should we report a single value for the LR or in addition address its uncertainty?

The heart of our different points of view concerns the way that parameter uncertainty[1] is dealt with, and can be found in section 4. Here, TBBA give an example of calculating the probability γ that an individual drawn at random from the relevant population has a particular DNA profile. TBBA consider

[1] There are also other sources of uncertainty that are of equal importance but are not discussed here nor in TBBA, such as model uncertainty.

this probability as a 'personal state of mind' that has no 'true' value, and hence cannot be 'estimated', it can only be 'assigned'. We agree with this because at this point, we have not yet defined the hypotheses of interest (H_p and H_d), the evidence (E), the relevant background information (I) and data (D), nor the probability models that we think generated the evidence under the two hypotheses. These definitions are all personal choices that can be made differently and hence γ has no 'true' value.

However, it is important to realize that TBBA nor we allow any arbitrary value to be assigned to γ. Indeed, when we specify H_p and H_d, E, I and suitable probability models, the value we must assign follows from probability calculus in the form of some expression containing certain parameters. These parameters are 'states of nature' that have a 'true' value which we can estimate based on the data D and prior probability densities (in the Bayesian paradigm). In TBBA's simple example, where e.g. population substructure is ignored, the value we must assign to γ is the population limiting frequency θ of the observed genetic profile in the population. Thus, at the moment we specify all elements of the model, γ is no longer a 'personal state of mind' but instead is equated to θ that has a true value.

Now if θ is known, e.g. 10%, there is no difference of opinion whatsoever. We agree with TBBA that γ is a single number with no uncertainty attached: it is simply 10% and this is the single number that should be reported to the legal justice system. However, in reality θ is not known, and we base our assessment on available information such as data sets. Following the example and notation of TBBA, suppose that we assume a prior beta distribution Be(α,β) for θ, and update our prior beliefs with a data set D of n random individuals from the population, where s individuals have the particular profile. We then obtain a known posterior Be($\alpha+s$,$\beta+n$-s) probability density[2] for θ. How should such information be reported? It is at this point where an important choice is made and our opinions diverge:

- *TBBA choose to inform the court about γ.* This probability can be obtained by integrating overall possible values of θ, weighting with its density according to probability calculus, and obtain the (posterior) expected value of θ, E(θ), which equals $\frac{\alpha+s}{\alpha+\beta+n}$. This is indeed a single known number with no uncertainty attached.
- *We choose to inform the court about θ.* Now θ is a population limiting frequency that has an unknown 'true' value, and there is uncertainty attached to it in the form of a known probability density Be($\alpha+s$,$\beta+n$-s). Hence, we can estimate θ by e.g. its mean $E(\theta)$ and report this to the legal justice system. However, as with any estimate we think it is also important to provide a measure of uncertainty in the form of an interval, e.g. a 95% credible interval.

Others (e.g. Brümmer and Swart, 2014; Cereda, 2015a,b) have described the choice between focus on γ and integrating versus focus on θ without integrating as choosing between including or excluding the data set D as part of the evidence to be evaluated[3]. In formulas: choosing between defining the LR as Pr(E, D|H_p, I]/ Pr(E, D|H_d, I] and integrating out all uncertainty, or as Pr(E|H_p, I]/ Pr(E|H_d, I] without integration. The first choice is called the '(full) Bayesian' approach, the second choice has different variants such as 'posterior LR'/'posterior Bayes Factor' (Aitkin, 2010; van den Hout and Alberink, 2015) when the data set D is used to obtain the posterior distribution of θ (as in the example above); 'plug-in' when D is used to produce point estimates of the parameters, resulting in a single value for θ

[2] Note that we follow TBBA's formula for their simplified example where e.g. the information that the suspect has the profile is not used. See Cereda 2015a for a discussion.

[3] As explained by Cereda, this is equivalent to choosing between including the data set D as part of the background information or not, since Pr[D|H_p, I] = Pr[D|H_d, I]. We thus choose between defining the LR as Pr(E|H_p, I, D]/ Pr(E|H_d, I, D] (with integration) or as Pr(E|H_p, I]/ Pr(E|H_d, I] (without integration).

that ignores parameter uncertainty; and ‘frequentist’ when D is used to produce a (frequentist) point estimate of θ and its sampling distribution to produce e.g. a confidence interval.

Hence, the question is: what is the most suitable choice when reporting to the legal justice system? Ultimately TBBA propose to report only $E(\theta)$, whereas we propose to report $E(\theta)$ (or some other estimate of θ, like the median of the posterior distribution) accompanied by an upper limit or interval. Thus, we provide more information than TBBA. We continue by arguing that the additional information we propose is important to the legal justice system.

3.2 *Why reporting the uncertainty of $E(\theta)$ is important for the legal justice system*

From a philosophical point of view, we believe that when the legal justice system asks an expert for his professional opinion, they are more interested in his opinion about ‘states of nature’ than in the expert’s personal ‘state of mind’. Translated to the example above, suppose that the legal justice system is given a choice between two types of information:

- The single number $E(\theta)$, representing the subjective degree of belief that the expert assigned to the event that a random person has the DNA profile.
- All the expert knows about θ, the limiting frequency of the DNA profile in the population: not only its estimated value $E(\theta)$, but also other values that the expert thinks are reasonably supported by the available data.

We reason below that the legal justice system should be fully informed about all possible values of the unknown parameter θ that in TBBA’s example completely determines the strength of the evidence, rather than be only informed about the single value that the expert assigns to this strength. In other words, the legal justice system is interested in ‘*all’* values that are supported by the data, not just in the ‘expected’ value.

The main reason that the legal justice system needs the full information is that they do not only need to assess the strength of the evidence, but also the reliability of the evidence. Based on the assessment of the reliability, important decisions are made concerning e.g. the need for additional research, counter-expertise, or even dismissal of the evidence. Suppose that we have a very partial DNA profile, and compare three experts:

(1) Expert 1 has profiled all individuals in the population, he is certain that θ is 10%.
(2) Expert 2 assumes, based on the very limited number of loci a Be(1,9) prior and profiles a thousand individuals: one hundred of them have the profile.
(3) Expert 3 also assumes a Be(1,9) prior and profiles 10 individuals: one of them has the profile.

Now when all experts go to court, they will all report exactly the same when they follow TBBA’s advice: report only $E(\theta)$. In this example, they will all report a probability of 10% that a random person has the profile. However, the expert opinions obviously are not equally reliable. This is hidden from the court when all experts report the same, and this could be considered as misleading. When the experts follow our advice, it would become clear that Expert 1 is the most reliable expert and Expert 3 the least reliable. Of course, one may argue that this is a special situation in which the expert’s prior expectation of θ equals the sample frequency (both are 10%). However, since $E(\theta)$ is a weighted average of the prior expectation and the sample frequency, the three experts will report a similar number whenever they are good experts whose prior expectations are close to the true value of θ. When different formulas than those of TBBA are used, we can adapt the example such that we still have three experts with the same expected values but different variances. Apart from different experts we can also think of different situations with the same expected values but with different variances. TBBA's final remark in section 6 about precision and rather vague suggestion to avoid numbers when communicating LRs will not solve the problem. We conclude that TBBA's advice to report only $E(\theta)$ deprives the legal justice system of essential information needed to assess the reliability of the evidence.

Another reason why the legal justice system may need the full information about θ and not just its expected value, is that they may want to give the benefit of the doubt to one of the parties involved, e.g. the suspect. Similar arguments are that they may choose to be 'on the safe side', or be 'conservative'. For this reason, quantifying uncertainty is a standard requirement for measurements on alcohol levels in blood or breath, or speed of vehicles. Furthermore, it is a standard topic in all modern quality assurance documents. Why would we consider this of the utmost importance in such measurements, but not in measurements of evidential strength? Only when we assume that judges maximize their expected utility in a decision framework, $E(\theta)$ is sufficient. A judge who wishes to be 'conservative' could in principle achieve this by choosing his utility functions, accordingly. However, whether judges follow this framework in practice is the question. We conclude that TBBA's advice to report only $E(\theta)$ and not its uncertainty may deprive the legal justice system of essential information needed when they do not work according to the rules of rational decision theory.

Interestingly, TBBA raise the question of reliability in section 5 for a LR of 10 reported in a specific case: 'Such a result may give rise to questions of the following kind: "How much credit should we owe to this result?", "To what extent can we rely upon this result?" '. These are very relevant questions, but unfortunately TBBA do not answer them, because they move on to explain why certain related approaches do not answer these questions. They conclude the section by stating that 'for a given case assessment, a comparison is evaluated in terms of a single value of likelihood ratio or Bayes factor'. Clearly, this does not provide an answer to the questions about reliability raised. The single value reported is the combined value of the forensic evidence E and the data set D. As a consequence, two experts with different databases will state different numbers for the evidential value, yet neither number has any uncertainty. This may be confusing and is another reason to prefer the second choice which focuses on the value of the evidence E only (Alberink *et al*., 2013).

3.3 *Estimation of an LR is not an abuse of language*

We agree with TBBA that it does not make sense to estimate something that has no true value. We also think TBBA would agree with us that it does make sense to estimate an expression of unknown parameters that all have 'true' values, such as population frequencies. As explained above, the LR transforms into such an expression by modelling the situation and defining all elements of it. In the DNA example, this expression boils down to θ. When we include the data set as part of the evidence to be evaluated, and subsequently integrate, we obtain a single number $E(\theta)$ where the variance is lost; we agree with TBBA that this number has no uncertainty and it makes no sense to talk about estimating it.

However, as explained above the integration is a choice. Why would we opt for this loss of information? We have argued above that it is much more informative for the legal justice system to explore the resulting parametric expression, which represents the true but unknown strength of the evidence only, and see what LR values we obtain for varying parameter values. In other words, instead of integrating out all parameter uncertainty and calculating the value of E and D combined, we can also explore the consequence of the uncertainty caused by limitations of D for the true but unknown strength of the evidence E. This means that a probability distribution of the LR (as a parametric expression) exists, and it is perfectly sensible (and not an abuse of language) to explore it, and estimate its mean or median and variance.

3.4 *Other proponents of studying uncertainty of an LR*

Despite the limited number of references in TBBA, we are not the only proponents of this line of thought in forensic science. Many other authors have also studied the uncertainty of an LR, e.g. Ali *et al*., 2015; Best *et al*., 2013; Beecham and Weir, 2011; Cereda, 2015b; Curran, 2005; Curran and Buckleton, 2011; Hancock *et al*., 2012; Morrison, 2011a,b; Morrison *et al*., 2011; Nordgaard and Höglund, 2011; Taylor *et al*., 2014; . Remarkably, the first author of TBBA could also have referred to his own publication to criticize (Dujourdy *et al*., 2003). Apparently, he has changed his mind about this topic.

4. Conclusion

In their paper, TBBA suggest that there is only one way to deal with parameter uncertainty that is mathematically correct and logically sound. In fact, however, there is a choice between a 'full-Bayesian' and a 'plug-in/frequentist/posterior LR' approach. The 'abuse of language' and 'misconception of principles' that TBBA aim to correct is merely a consequence of choosing the latter approach, and for this approach the principles and language used are perfectly suitable. There is no mathematical or logical reason to prefer one approach over the other: both have pros and cons. However, for application in the forensic domain, the 'full-Bayesian' approach advocated by TBBA deprives the legal justice system of essential information. We conclude that forensic scientists should not only report an LR value, but also address its uncertainty.

References

Aitkin M. (2010) Statistical Inference. *An Integrated Bayesian/Likelihood Approach*, Chapman & Hall/CRC, Boca Raton, FL.

Alberink I., Bolck A. and Menges S. (2013) Posterior likelihood ratios for evaluation of forensic trace evidence given a two-level model on the data, *Journal of Applied Statistics* **40**, 2579–2600.

Ali T., Spreeuwers L., Veldhuis R. and Meuwly D. (2015) Sampling variability in forensic likelihood-ratio computation: A simulation study, *Science & Justice* http://dx.doi.org/10.1016/j.scijus.2015.05.003.

Beecham G. W. and Weir B. S. (2011) Confidence Interval of the Likelihood Ratio Associated with Mixed Stain DNA Evidence, *Journal of Forensic Sciences* **56**, S166–S171.

Best N., Ashby D., Dunstan F., Foreman D. and McIntosh N. (2013) A Bayesian approach to complex clinical diagnoses, a case-study in child abuse, *Journal of the Royal Statistical Society: Series* A **176**, Part 1, pp. 53–96.

Brümmer N. and Swart A. (2014) Bayesian calibration for forensic evidence reporting, arXiv:1403.5997v3 [stat.ML] 10 Jun 2014.

Cereda G. (2015a) Impact of model choice on LR assessment in case of rare haplotype match, (Bayesian approach) arXiv:1502.02406v7 [stat.AP] 24 Jun 2015.

Cereda G. (2015b) Impact of model choice on LR assessment in case of rare haplotype match, (frequentist approach) arXiv:1502.04083v5 [stat.AP] 24 Jun 2015.

Curran J. M. (2005) An introduction to Bayesian credible intervals for sampling error in DNA profiles, *Law, Probability and Risk* **4**, 115–126.

CURRAN J. M. AND BUCKLETON J. S. (2011) An investigation into the performance of methods for adjusting for sampling uncertainty in DNA likelihood ratio calculations, *Forensic Science International: Genetics* **5**, 512–516.
DUJOURDY L., BARBATI G., TARONI F., GUNIAT O., ESSEIVA P., ANGLADA F., AND MARGOT P. (2003) Evaluation of links in heroin seizures, *Forensic Science International* **131**, 171–183.
HANCOCK S., MORGAN-SMITH R. AND BUCKLETON J. (2012) The interpretation of shoeprint comparison class correspondences, *Science & Justice* **52**, 243–248.
VAN DEN HOUT A. AND ALBERINK I. (2015) Posterior distributions for likelihood ratios in forensic science, unpublished manuscript, available since 2015 at http://www.homepages.ucl.ac.uk/~ucakadl/.
MORRISON G. S. (2011a) A comparison of procedures for the calculation of forensic likelihood ratios from acoustic-phonetic data: Multivariate kernel density (MVKD) versus Gaussian mixture model-universal background model (GMM-UBM), *Speech Communication*, **53**, 242–256.
MORRISON G. S. (2011b) Measuring the validity and reliability of forensic likelihood-ratio systems. *Science & Justice* **51**, 91–98.
MORRISON G. S., ZHANG C. AND ROSE P. (2011) An empirical estimate of the precision of likelihood ratios from a forensic-voice-comparison system. *Forensic Science International* **208**, 59–65.
NORDGAARD A. AND HÖGLUND T. (2011) Assessment of approximate likelihood ratios from continuous distributions: A case study of digital camera identification, *Journal of Forensic Sciences* **56**, 390–402.
TARONI F., BOZZA S., BIEDERMANN A., AND AITKEN C. (2016) Dismissal of the illusion of uncertainty in the assessment of a likelihood ratio. *Law, Probability and Risk* **15**, 1–16.
TAYLOR D., BRIGHT J.-A., BUCKLETON J. CURRAN J. (2014) An illustration of the effect of various sources of uncertainty on DNA likelihood ratio calculations, *Forensic Science International: Genetics* **11**, 56–63.