Taylor, Duncan, Bas Kokshoorn, and Christophe Champod. "A Practical Treatment of Sensitivity Analyses in Activity Level Evaluations." Forensic Science International 355 (2024/02/01/ 2024): 111944.

KEY WORDS:

Activity level; Bayesian Network; sensitivity analysis; court reporting; DNA profile evaluation.

HIGHLIGHTS:

- Forms of sensitivity analysis that apply to activity level evaluations are presented
- Examples are given for when evaluations are likely to be most sensitive to data
- Strategies for practically dealing with sensitivity evaluations are discussed
- Examples of reporting the results of sensitivity evaluations are provided

ABSTRACT:

Evaluations of forensic observations considering activity level propositions are becoming more common place in forensic institutions. A measure that can be taken to interrogate the evaluation for robustness is called sensitivity analysis. A sensitivity analysis explores the sensitivity of the evaluation to the data used when assigning probabilities, or to the level of uncertainty surrounding a probability assignment, or to the choice of various assumptions within the model. There have been a number of publications that describe sensitivity analysis in technical terms, and demonstrate their use, but limited literature on how that theory can be applied in practice. In this work we provide some simplified examples of how sensitivity analyses can be carried out, when they are likely to show that the evaluation is sensitive to underlying data, knowledge or assumptions, how to interpret the results of sensitivity analysis, and how the outcome can be reported. We also provide access to an application to conduct sensitivity analysis.

1.0 INTRODUCTION:

Evaluation of forensic observations is recommended to occur using the likelihood ratio (LR) [1-5], whereby competing propositions are set to align with the positions of the prosecution and a viable alternative (often referred to as the defence proposition) [6]. The proposition can be set at different levels within the hierarchy of propositions [7, 8] being typically for DNA findings at sub-sub-source, sub-source, source, activity, or offence. When an evaluation of observation occurs,

regardless of the level which the propositions address, there is uncertainty about the state of the world in which the evaluation exists. A common example is that when evaluating DNA profiles considering propositions at the sub-source level, the frequency of alleles in the population is required. These are obtained referring to population databases, which are usually a small subset of the entire population. The database is assumed (due to underlying assumptions around the randomness of sampling individuals) to be representative of the population, however there is some uncertainty as to how representative it truly is.

It is common in forensic science to model the effect that the uncertainty can have on the evaluation. One such mechanism for doing so is called a sensitivity analysis. In essence the purpose of a sensitivity analysis is to determine how sensitive the LR is to the underlying data. Combined with the level of informedness that this data provides about the real world, a decision can be made as to whether the evaluation is robust. There have been numerous publications that seek to address how best to utilise or communicate measures of uncertainty (or precision) associated with a LR such as in the special issue of the journal Science & Justice 2016, volume 56, issue 5, [9-16], preceded by a discussion in Law, Probability & Risk [17, 18]).

Despite all this discussion there is limited practical guidance on how to utilise the results of a sensitivity analysis. For example, when is the LR so sensitive that the robustness of the likelihood ratio comes into question? If this does occur, what should be communicated to stakeholders? Most scientists would agree that if a LR is very sensitive to one particular source of underlying data, and the amount of data available in that source is very low then it is likely that the LR is not robust (noting that we deliberately do not define 'very sensitive' or 'very low'). This point was brought up in [16], where the authors showed two examples (in Figure 9 of their paper) of an evaluation leading to an LR ~ 1, but one that was highly sensitive to the underlying data and the other insensitive to the underlying data. Taylor et al. [19] later suggest one mechanism for dealing with the results of a sensitivity analysis is that once carried out the analyst can choose conservative parameter values and report the resulting LR (where conservative here is meant to mean that it minimised the strength of evidence that the LR portrays).

Sensitivity analysis methods

There have been different methods explored for carrying out sensitivity analyses, which are explained in [20]. In summary, the simplest method can be carried out when dealing with a binary

system (i.e., in a locus with only two alleles, or when considering the probability that transfer did, or did not, occur in a particular scenario). In this situation the probability assigned to one state can simply be varied across the range (0,1), with the probability assigned to the other state deterministically defined as the complement of that for the first state. The value of the LR is then calculated for each value assigned to the underlying probability. An example of this type of sensitivity analysis was shown in [21]. This method for carrying out sensitivity analysis can examine one factor (or its assigned probability) at a time to show the sensitivity of the LR to each probability between 0 and 1. The idea can be extended to consider the effects of two factors at once (see Figure 2 in [21]), or this could equally be applied to a factor that possessed three states (e.g., a locus with three alleles, or a transfer that has probabilities assigned for states high, low, none). Going beyond this level of complexity takes the calculation of the sensitivity of the LR to specific states within a factor beyond reasonable ability to visualise. When this occurs the sensitivity analysis is carried out by simulated resampling of the underlying data and the resulting values usually displayed in a boxplot showing the spread of LRs obtained (as seen in [16]). Up until this point the sensitivity analyses described are used when probability assignments are based on counts of observations that fall into two or more states. When the states of factors within the evaluation are set by modelling the distribution of underlying data then the sensitivity analysis must accordingly simulate the data and then the modelling in order to obtain sensitivity information [22].

LRs and robustness

Within the context of evaluations considering sub-source level propositions, it is common to define a desired reproducibility of within one order of magnitude [23, 24]. Bright et al. [23] state:

> *The variability within the central 95% of sub-source point estimate LRs (with $\theta = 0.01$) was found to be within the expected one order of magnitude*

This comes as some probabilistic genotyping systems are based on stochastic sampling processes (such as Markov Chain Monte Carlo) and so each time an analysis is run it will lead to a different result (with the exception of when the stochastic process is deliberately fixed to take the same path). This repeated analysis tests the sensitivity of the LR to the statistical model. Note that this is different to the resampling of the LR (for example using the highest posterior density, HPD, method [25]), which is also a form of sensitivity analysis. A HPD analysis can also encompass the

variability of the statistical model, but also other aspects such as allele frequency sampling variation deriving from the construction of a population database, or the uncertainty around the level of co-ancestry within a population [26]. While a very wide distribution of LRs resulting from a HPD analysis can pose an issue, there is no specific guidance on how wide the distribution can be before an analysis is deemed to no longer be robust.

Within the context of evaluations considering activity level propositions Samie et al. [27] define 'high variability' by:

> *A Likelihood Ratio (LR) is obtained for each simulation and the posterior probability is calculated from this LR and the selected prior probabilities* [text removed]. *For our laboratory, we have arbitrarily defined that the variation of the posterior probability is high when the ratio interquartile to the median is higher than 0.2.*

The acceptable level of variability on the LR given the definition of Samie et al [27] quickly broadens as the LR increases. The authors make the point that the method they employ is only applicable to the specific situation they are applying it to, and so would not be applicable as a general method of assessing when the LR is sensitive to some underlying data.

In our work we consider a simple method and definition on how an evaluation can be interrogated for robustness. By necessity the definitions we provide are arbitrary, but we explain them in a way that would allow another laboratory to apply the same thinking but choose different definitions for 'sensitive'. We mainly focus on the sensitivity analysis in the context of activity level evaluations, and the use of Bayesian Networks (BN). Therefore, we consider different factors in an evaluation that equate to nodes in a BN, and the various states that the node can take. We also provide guidance on how to deal with the issue of an LR that is sensitive to underlying data. Finally, we provide an online tool to assist in carrying out sensitivity analyses from Bayesian network files.

The aim of this work is to provide readers with a greater understanding of sensitivity analyses, and translate the often complex published theory into practically applicable actions. We also aim to take the first steps towards providing a free online resource that allows users to carry out the sensitivity analyses demonstrated in this paper themselves, without the need for programming knowledge.

2.0 METHOD:

Consider a slightly modified scenario from Taylor et al. [28]:

> *A suspect (Mr. S) attends the same party as the victim (Ms. V) and is accused of digitally sexually assaulting the victim at some point during the night. Swabs of Mr S's fingernails are taken and reveal a DNA profile that has the same characteristics as the combined profiles of Mr S and Ms V. Mr S denies the digital assault but agrees that he socially interacted with Ms V at the party.*

A simplifying assumption is made that the source of the DNA is thus not in dispute and so only the disputed activities are considered.

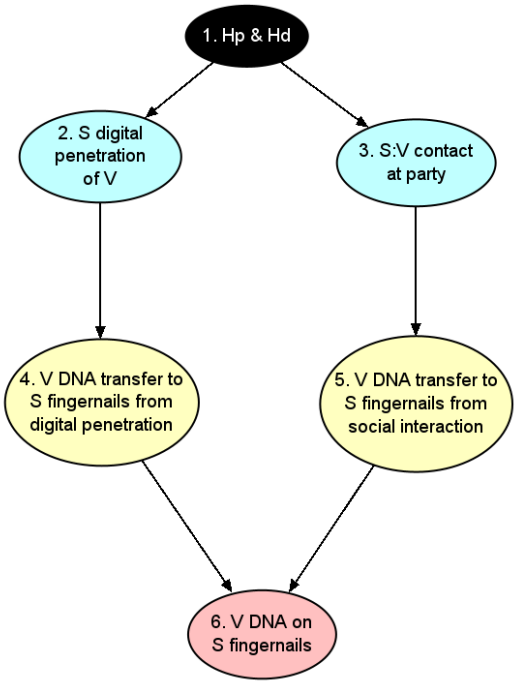
The propositions are:

H_p) Mr S inserted his fingers in Ms V's vagina.

H_d) No one inserted their fingers in Ms V's vagina

Background: I) Mr S and Ms V socially interacted that night

In Taylor et al. [28] they provide a simple BN (constructed using the BN template of [29]), which we redraw a slightly modified version of in Figure 1. The BN was developed using the BN commercial software Hugin (www.hugin.com). We also recreated the BN using software R [30] and the freely available R libraries, gRain [31] and BNlearn [32]. Note that BN prepared in Hugin and saved as .net file can be directly imported as gRain or BNlearn objects.



1. Hp & Hd		
	Hp	0.5
	Hd	0.5

1. Hp & Hd		Hp	Hd
2. S digital penetration of V	Yes	1	0
	No	0	1

1. Hp & Hd		Hp	Hd
3. S:V contact at party	Yes	1	1
	No	0	0

2. S digital penetration of V		Yes	No
4. V DNA transfer to S hand from digital penetration	Yes	0.7	0
	No	0.3	1

3. S:V contact at party		Yes	No
5. V DNA transfer to S hand from social interaction	Yes	0.18	0
	No	0.82	1

4. V DNA transfer to S hand from digital penetration		Yes		No	
5. V DNA transfer to S hand from social interaction		Yes	No	Yes	No
6. V DNA on S hand	Yes	1	1	1	0
	No	0	0	0	1

Figure 1: BN constructed for evaluation of DNA results given sexual assault scenario explained in the main text (left) and associated probability tables (right). In the probability tables the nodes to which the individual tables apply have been bolded.

Nodes 1, 2, 3 and 6 in the BN in Figure 1 are assigned probabilities based on proposition prior probabilities, case circumstances or are logically determined by their parents. In each case the probability does not rely on experimental data. For example, the case circumstances dictate whether activities have occurred under Hp, Hd or both and this naturally translates to assigning 0 or 1 in the cells of tables for nodes 2 and 3 (without needing any experimental data). similarly, the presence of V's DNA on the hand of S will occur if either DNA has come to be there from digital penetration or social interaction, and again this leads naturally to assigning 0 or 1 in the cells of the table for node 6 without the need for any experimental data.

Probabilities for nodes 4 and 5 in Figure 1 rely on counts of experimental observations. As per [16] we apply a hyperprior vector of 1's in a prior Dir(1,…,1) and use it to calculate the posterior mean probability of each count for use in the BN. Each node has a conditional probability table in which a column represents a combination of parental factors referred to as a category. Each row represents a value that the node can take and is referred to as a state. If state *i* of category *k* has

157 $n_{i,k}$ observations, then the Dirichlet distribution that describes the multi-dimensional probabilities
158 is:

159 $Dir\left(n_{1,k+1},...,n_{I,k+1}\right)$ where I is the number of different states that exist in that category

160 and the posterior probabilities are calculated by:

161 $p_{i,k} = \frac{n_{i,k}+1}{I+N_k}$ where N is the total number of observations in category k, $N_k = \sum_i n_{i,k}$

162

163 Node 4 in Figure 1 considers the recovery of DNA of the vaginal cavity from hands after digital
164 penetration. We use the work of Flanagan et al. [33] who looked at DNA persistence under
165 fingernails over different time frames following digital penetration. The appropriate timeframe to
166 use from the study is that which aligns most closely with the circumstances of the case. For the
167 worked example we use the data for 18 hours, which was the longest period investigated. Only
168 two out of eight fingernail samples did *not* detect the female partner's DNA after 18 hours.
169 Probabilities of (6+1)/(8+2) ≈ 0.7 is assigned for recovery of DNA and (2+1)/(8+2) ≈ 0.3 for no
170 recovery. If digital penetration did not occur, then DNA cannot transfer through this mechanism
171 so probabilities of 0 for yes and 1 for no are assigned.

172 Node 5 in Figure 1 considers the probability of DNA being recovered from the fingernails of the
173 defendant after a social interaction with the victim. Goray et al. [34] conducted DNA transfer
174 experiments within a social setting. In each of their 4 experiments there were 3 participants, who
175 had 2 hand swabs taken (one from each hand) taken at the conclusion of a social interaction. This
176 provides 4 (experiments) x 2 (hand swabs) x 6 (combinations of participant transfer) = 48
177 opportunities for a participant's DNA to be recovered from another participant's hand. Of these
178 48 events, there were 8 observed transfers. Probabilities of (8+1)/(48+2) ≈ 0.18 is assigned for
179 recovery and (40+1)/(48+2) ≈ 0.82 for no recovery.

180 As the assignment of probabilities themselves is not the central issue of this paper we will assume
181 contestably that the experimental setup of the work carried out in Goray et al. [34] and Flanagan
182 et al. [33] align closely with the circumstances of the case. Under this assumption, the uncertainty
183 which exists is from the study size, and not due to the conditions in the study diverging from the
184 circumstances of the case, there being unknown factors about the circumstances of the case, or the

analytical methods used in the study being different from those in the investigating laboratory. We accept this is a narrow definition under which to conduct sensitivity analyses, but it represents a starting point from which practical guidance can be developed. We will expand on this assumption within the discussion.

If the BN in Figure 1 is provided with the information that V's DNA is present on S's fingernails, then the LR obtained is approximately 4.19 i.e., the probability of observing V's DNA on S's fingernails is about 4 times higher if he digitally penetrated V's vagina, than if he didn't. This can be demonstrated by deriving the formula. If we define:

$\Pr(T_D)$ - the probability of transfer from the vagina of the complainant to fingernails of the defendant from digital penetration

$\Pr(T_S)$ - the probability of transfer from the complainant to fingernails of the defendant from social interaction

then

$$\Pr(E \mid H_p) = \Pr(T_D)\Pr(T_S) + \Pr(T_D)\Pr(\overline{T_S}) + \Pr(\overline{T_D})\Pr(T_S)$$

$$\Pr(E \mid H_d) = \Pr(T_S)$$

where the lines over the top of the terms represent a non-transfer and are the complement of the transfer probability i.e., $\Pr(T.) = 1 - \Pr(\overline{T.})$. With some simplification, the LR is:

$$\mathrm{LR} = \frac{1 + \Pr(\overline{T_D})\left[\Pr(T_S) - 1\right]}{\Pr(T_S)}$$

Taking the probabilities assigned based on the two studies we arrive at an LR of 4.19, the same value as given by the Bayesian Network.

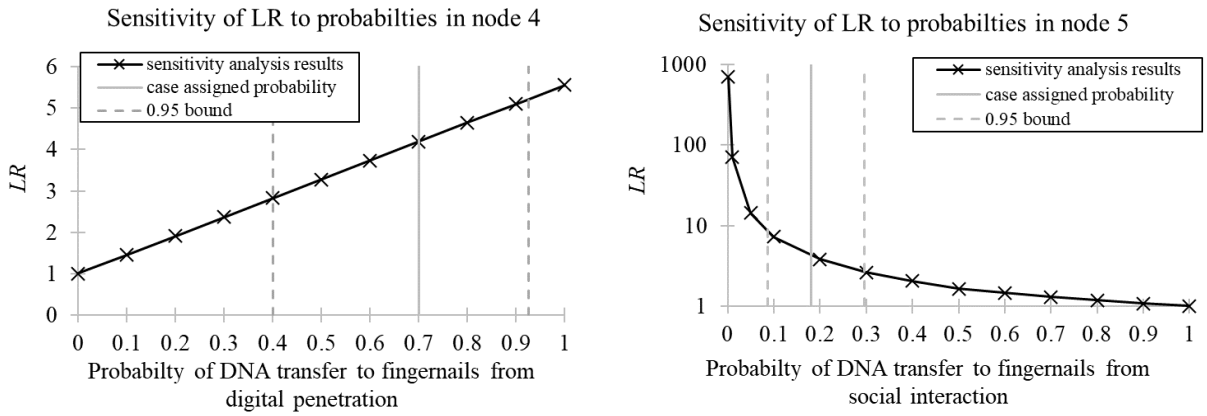
3.1 - Sensitivity analyses on a binary state node – with paired data

In this section we consider the situation where data is available and believed to appropriately represent the circumstances of the case to which it is being applied. We refer to this as 'paired data'. In this situation, the uncertainty in the evaluation comes purely from the size of the study. We later discuss the situation where data does not exactly represent the scenario in the case or is not available.

Start by considering a classic sensitivity analysis where the probabilities assigned in nodes 4 and 5 are varied across the entire possible range [0, 1]. The resulting effect on the LR is shown in Figure 2. When there are only two states in a factor (i.e., transfer and no-transfer) then the Dirichlet distribution used to model counts is reduced to two parameters, which is a beta distribution. Therefore, in the simple case of a binary node in a BN the beta distribution can be used to determine the desired interval given by a sensitivity analysis. Consider the results of the transfer experiment of Flanagan et al. [33] used in node 4. Out of eight experiments there were six observations of transfer. A Beta distribution has two shape parameters, α and β, and can be written as $B(\alpha, \beta)$. If we define N_k as the number of experiments and $n_{t,k}$ as the number of transfers, then using a B(1,1) prior the posterior probability of DNA transfer can be modelled by

$$B\left(n_{t,k}+1, N_k - n_{t,k}+1\right)$$

From this distribution the desired bounds can be chosen. For example, the two-sided 95% bounds, which corresponds with the central 0.95 of the area under the Beta distribution use to model transfer probability. The lower and upper bound values for the central 0.95 area are the values corresponding to 0.025 and 0.975 of the area of the Beta distribution. These lower and upper bound values can then be used as the bounds on the LR sensitivity analysis. In Figure 2 we show the effect of changing transfer probability on the LR, with the lower and upper bounds on transfer probability corresponding to the two-sided 95% bounds of the Beta distribution used to model it.



232 *Figure 2: Sensitivity analysis for the BN shown in Figure 2 to data underlying nodes 4 (left) and*
233 *5 (right).*

234

235 The sensitivity analysis results shown in Figure 2 can be obtained by manipulation of the
236 probabilities in the BN, or in simple cases (such as this example) the LR formula can be derived
237 and used directly as shown in the previous section.

238

239 Using this formula (or manipulation of the BN, however this becomes somewhat tedious when
240 multiple nodes are involved) a plot can be produced that displays the sensitivity of the LR to both
241 factors at once, as shown in Figure 3.

242

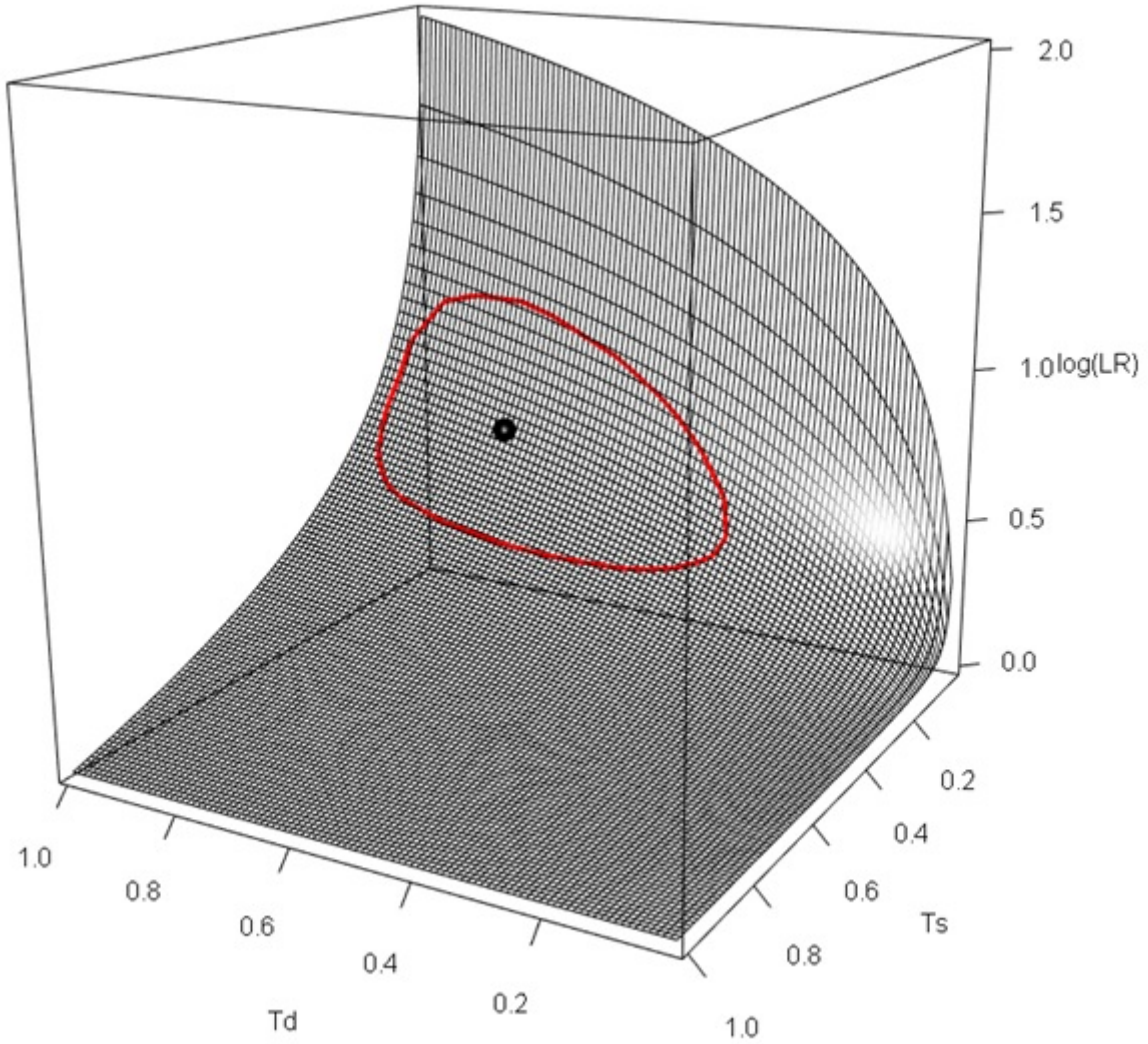


Figure 3: sensitivity of LR derived from BN in Figure 2 for data used in nodes 4 and 5 with 95% bound showing in red and case value indicated with a black point.

In Figure 3 we show a surface that represents the $\log_{10}(LR)$ value for each combination of $\Pr(T_D)$ and Pr(Ts) (broken into 200 steps between 0 and 1). The black point on the surface represents the *LR* using the case assigned probabilities and the red shape shows the 95% bounds based on the underlying data (in the same way as the grey dashed lines were shown in Figure 2, but now for this three-dimensional representation).

3.2 - Sensitivity analyses on multi-state nodes, or multiple nodes – paired data

Again, we still consider the situation in which we accept the circumstances of the study from which the data is being used, relates to the circumstances of the case. Therefore, we are interested in how sensitive the *LR* is to the number of experiments conducted. In section 3.1 the bounds could be set that represented a desired interval based on a beta distribution. We are now interested in data with higher dimensionality and so need the multinomial equivalent distribution. This distribution is the Dirichlet distribution, which can extend to any number of dimensions (including 2 in which case it simplifies back to the standard beta).

Many programs have functions for drawing values from standard distributions such as normal, gamma or beta distribution. A sample can be drawn from a Dirichlet distribution by utilising the gamma distribution in the following way:

Consider a factor that has I states (for example node 4 in the BN in Figure 1 has two states, Yes and No). Draw each observational count n_i by:

$n_i \sim gamma(n_i + 1, 1)$ for each n_i

and then normalise the I drawn values to obtain posterior probabilities, p_i:

$$p_i = \frac{n_i}{\sum_i n_i}$$

In order to carry out a sensitivity analysis for factor k, this process is carried out Y times to produce

$p_{i,k,1} \ldots p_{i,k,Y}$

These can be used to calculate Y *LR*s, which can then be ordered so that the values corresponding to the desired bounds can be obtained. This is carried out for each of the nodes based on observational data.

While nodes 4 and 5 in the BN from Figure 1 are both binary, we carry out the described resampling on both nodes individually, and together to produce the results seen in Figure 4. To obtain the bounds is then a matter of sorting the data and choosing the quantiles of interest i.e., for the 95% bounds the 0.025 and 0.975 quantiles are chosen.

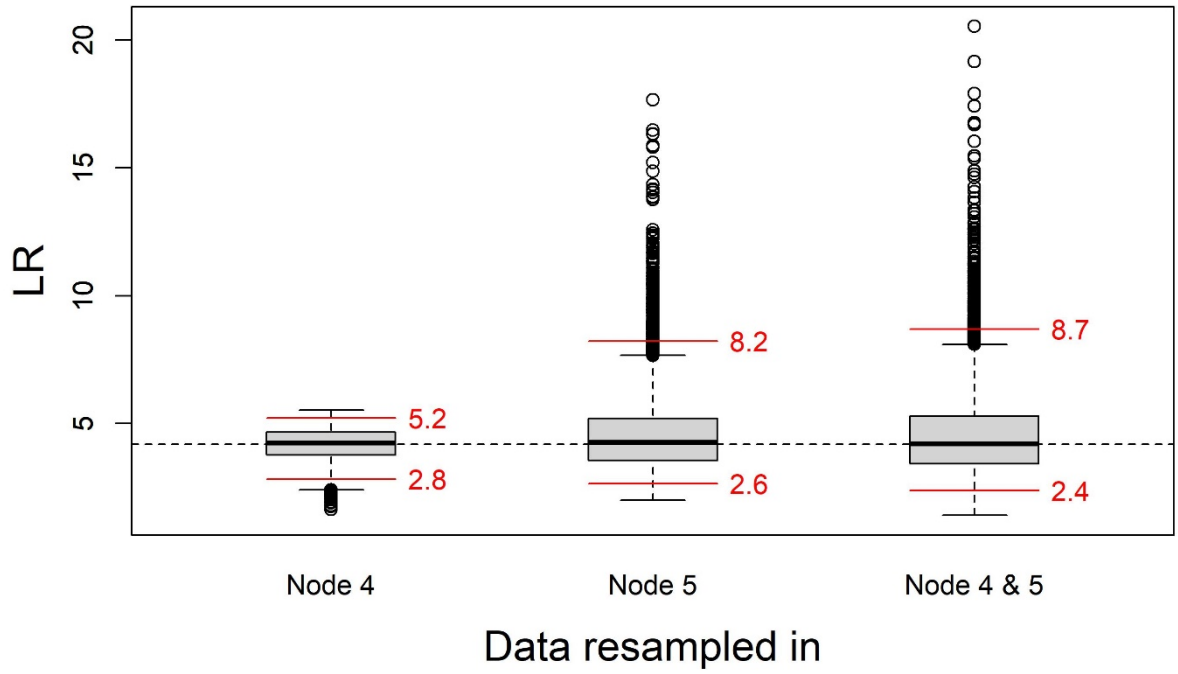


Figure 4: results of sensitivity analysis on the BN from Figure 1 using the resampling method (n=10000). The dashed line shows the LR obtained when using the case assigned probabilities. The red lines show the bounds relating to the central 95% quantile of the resampled dataset.

3.3 - When is sensitive too sensitive?

Consider the example of a sensitivity analysis that was shown in Figure 2. We may decide that the choice of designating an evaluation as sensitive will be based on the range of *LR* values obtained when varying some underlying probability. In Figure 2, we also demonstrate that the limits (or bounds) over which the underlying probability may be chosen to vary can be based on the bounds of the distribution being used to model it. This then translates to bounds over which the *LR* will be considered as varying in the sensitivity analysis (i.e., between the dashed grey lines in Figure 2). There still remains the need to interpret the result (i.e., if we take Figure 2 left then the bounds on the LR are 3 and 5, but is this robust?). For the purpose of this paper, the need to interpret the results of sensitivity analyses comes from a need to designate whether the *LR* is considered highly sensitive to some underlying data (or in other words, to designate when the evaluation is robust and can be reported). Arguments have been put forward that this choice is arbitrary and

unnecessary, and that the sensitivity analysis speaks for itself without further categorisations. In theory we agree with this sentiment, however in practice we argue that a designation is required. We save the details of our argument of this point of view to the discussion and ask that the reader accepts at this point that some results-based definition of "sensitive" compared to "robust" is needed.

The most obvious property of the sensitivity analysis results to use as a tool that can categorise an *LR* as highly sensitive is the relative difference either between the upper and lower bounds of the sensitivity analysis, or the mean *LR*[1] and one of the bounds. The exact value chosen for this threshold is arbitrary, as was the choice of desired bounds in section 3.1. Unlike the desired bounds (for which there is a convention of 0.95) there is little guidance on value to choose for a sensitivity threshold.

Consider first using the relative difference between the *LR*s corresponding to the upper and lower bounds of the sensitivity analysis. Using the difference between these values as an indication of sensitivity is likely to be overly conservative, because it is not usually a concern that the *LR* could have provided more support for a proposition than the reported value, but rather the greater concern is when the *LR* could have provided much less support (i.e., the main issue with sensitivity is not to overstate the strength of the observations). Therefore, a more sensible choice for categorising sensitivity maybe that the bound leading to the least discriminating *LR* is not more than some threshold from the mean.

A simple option for the choice of a threshold value could be to choose that when the difference in *LR* between the desired bounds are within an order of magnitude the evaluation is not highly sensitive. Another option, equally valid, could for instance be a factor of 2. The choice of threshold itself on what constitutes an evaluation as being highly sensitive is likely to be based, in part, on the experience of the laboratory and their interaction with stakeholders. We are not

[1] There is a technical point to note here that relates to the difference between the 'mean LR' and the LR obtained using the case derived posterior probabilities (often referred to as the 'point estimate'). The mean LR can be the point estimate (when only the one LR is available using the case derived posterior probabilities) but can also be the mean of the LRs obtained during sensitivity analysis simulation. When a point estimate is available, and a simulation has been carried out (so that a mean LR can be calculated from the simulation results) then the two values may not be the same. This comes from the fact that the LR is calculated as a ratio of probabilities. Take a very simplistic example where the LR is simply the ratio of probabilities of transfer under two different mechanisms. Consider the probability of transfer to have binary states (e.g., 'transfer' and 'no transfer') and so the probability of these transfers is each a Beta random variable, call them 'A' and 'B'. The point estimate of the LR is the ratio of the expected values for these variables, E[A]/E[B]. The mean LR from simulation is E[A/B], which does not necessarily equal E[A]/E[B]. When we refer to bounds being placed around a LR (such as an order of magnitude) we are referring to the mean LR from simulation, rather than the point estimate.

advocating any particular value in this work but will use an order of magnitude difference as an example, which will allow a number of further concepts to be demonstrated.

If this criterion is used to denote a highly sensitive *LR*, then in many instances the use of a hyperprior count vector (1,…,1) negates the greatest effects of sensitivity. For example, imagine that the results of the Goray et al [34] study were that transfer in a social setting was not observed. This would lead to a probability of $(0 + 1)/(48 + 2) \approx 0.02$ for transfer and 0.98 for no transfer, which leads to an $LR \approx 35$. The 95% bounds using a posterior B(1, 49) distribution are 0.000517 and 0.072519, which leads to *LR*s of 10 to 1355 i.e., the lower bound interval is only a factor of three from the reported value. Even if the number of experiments were 1000 the $LR \approx 701$ and the lower bound is 191, still less than an order of magnitude. In the other direction if there were only a single experiment carried out (again resulting in no transfer being observed) then the $LR \approx 2$ and the lower bound is 1.

3.4 - Sensitivity analyses – without paired data

There are going to be instances during casework evaluation where:

- There are differences between the circumstances of the case and the circumstances of the study from which data is being used (such as the laboratory techniques) and these must be accounted for by using a subjective adjustment of the probability that is assigned based on the frequencies obtained from the data.
- There are multiple studies that are being relied on and they result in different probabilities.
- There is a lack of applicable published data and a reasonable range for the probability is defined using secondary sources (i.e., studies that are not directly relevant, but could inform an upper or lower bound on what would be expected for the probability)
- The assignment for the probability is made subjectively using experience or knowledge, but without being informed by any identifiable studies.

In these situations, there are different ways in which the evaluation (and potentially sensitivity analysis) can proceed. There may be interest in how sensitive the *LR* is to the probability assigned to the factor.

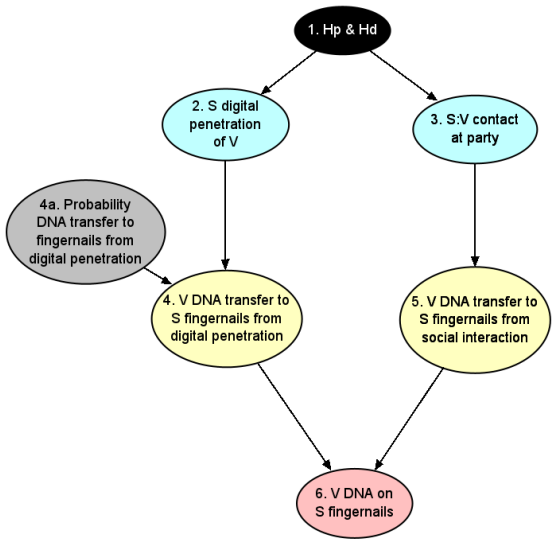
In this case the probability can be varied across the range that is deemed to be sensible (or even across the entire range (0,1) if the interest extends that far) and the resulting *LR*s examined (producing the same type of plot as seen in Figure 2, but with the bounds representing the subjectively defined bounds rather than 95% bounds).

When experience-based probabilities are assigned to a multi-state factor one possible solution would be to consider an experience-based assignment as equivalent to a data-based assignment of *N* samples (where *N* could take any reasonable value such as 5, or 10, or 100).

The key point of difference is that this is not carrying out a sensitivity analysis that is considering the effect of a dataset sample size even though, mechanically, the same calculation process is occurring. Instead, the sensitivity analysis is being carried out in order to investigate the effect of *a lack of* data available to inform a probability assignment. The counts introduced capture the postulated extent of the knowledge used to inform the conditional probability tables. For example, in the absence of the Goray et al [34] study, we could have assigned a probability of 0.02 for transfer and 0.98 for no transfer invoking conceptually a total of 50 "experiments" modelled with the same B(1, 49). Note that the "experiments" are the result of a thought experiment to best represent one's knowledge and experience. Using this method, the number of hypothetical experiments can be set to reflect the confidence the expert has in their probability assignment i.e., the higher the number of experiments, the greater the confidence.

If the sensitivity of the *LR* to the probability is not in and of itself a point of interest, then it may be that a sensitivity analysis is not the best way to deal with the uncertainty associated with experience-based probability assignments at all. Instead, if the aim of the sensitivity analysis is to assess the risk that the *LR* could be overstating (or understating, although this is usually of lesser concern) the strength of evidence then an experience-based probability assignment can be dealt with by assigning a prior distribution for the assignment that reflects the uncertainty held by the scientist. This may be done in a similar way as in a situation where we are dealing with multiple sources of equally (ir)relevant data. Consider node 4 in the BN in Figure 1 that accounts for the probability of DNA transfer to someone's hands from digital penetration of a vagina. The assignment shown in Figure 1 was 0.7 that the transfer would occur and 0.3 that it would not. Instead consider a situation where several studies had been found that showed probabilities for DNA transfer that ranged from 0.9 to 0.5 and *a priori* each appeared equally valid to apply to the case. One option would be to extend the BN so that node 4 had a parent node with states for the probability of transfer between 0.5 and 0.9 each with an equal prior probability (hence a uniform

distribution). Alternatively, the collection of studies could be considered as suggesting that a central value of 0.7 is appropriate, but some account of the fact that other values are possible. In that instance, again a parent node could be created with states that span 0.5 to 0.9 but with the probability distributed according to a normal distribution with a mean of 0.7 and a reasonable variance that spread the probability in a way that reflected the scientist's belief. Figure 5 shows an example of a BN set up in this way.



4a. Probability DNA transfer to fingernails from digital penetration	
0.5	0.2
0.6	0.2
0.7	0.2
0.8	0.2
0.9	0.2

2. S digital penetration of V		Yes					No				
4a. Probability DNA transfer to fingernails from digital		0.5	0.6	0.7	0.8	0.9	0.5	0.6	0.7	0.8	0.9
4. V DNA transfer to S fingernails from digital penetration	Yes	0.5	0.6	0.7	0.8	0.9	0	0	0	0	0
	No	0.5	0.4	0.3	0.2	0.1	1	1	1	1	1

Figure 5: BN constructed for evaluation of DNA results extending the BN from Figure 1 but including a parent node (4a) to reflect a prior distribution for the probability in node 4 (left) with additional node probability table (right). In the probability tables the nodes to which the individual tables apply have been bolded.

The advantage of expressing uncertainty in the fashion shown in Figure 5 is that in effect the uncertainty in the appropriate probability assignment is being integrated out of the evaluation [11]. There is no longer any need to carry out a sensitivity on this factor. There is also no need to carry out sensitivity analyses on the parameters of the parent node (i.e., the upper or lower bounds being considered, or the parameters of a distribution) as the parent node already represents our uncertainty (and it does not make sense to consider the uncertainty held over the level of uncertainty).

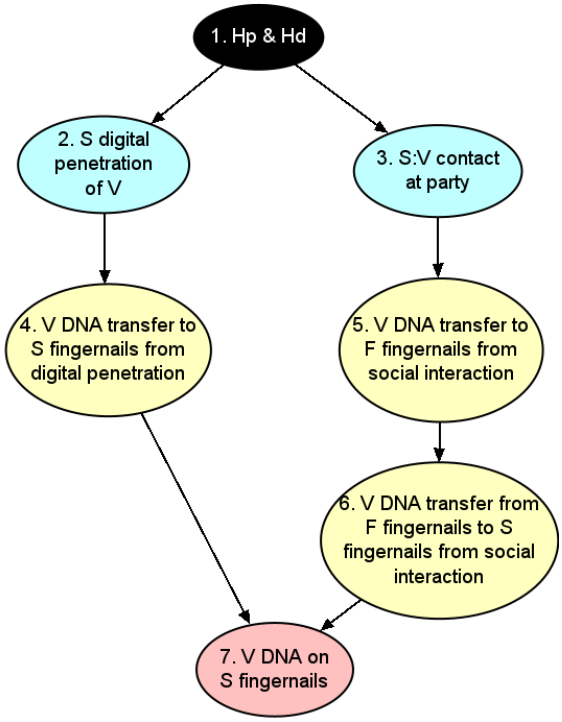
Note that even in situations with available paired data where a sensitivity analysis investigates the effect of sample size on the *LR* the method of integrating out the uncertainty can be conducted. Instead of drawing values from a beta distribution and identifying 95% bounds, the beta distribution itself could be expressed in a parent node. Again, this would remove the need for a sensitivity analysis on the effect of the limited data on the *LR*. If the *LR* was highly sensitive to the probability value at one end of the range (for example the exponential rise in *LR* at lower probability values for node 5 in the BN from Figure 2, shown graphed in Figure 2 right), then integration across the prior probably distribution will dampen any effect on the *LR*.

3.5 - When the *LR* may be highly sensitive

Given the demonstration of how using a hyperprior count vector (1,…,1) negates much of the issue of sensitivity of the LR to the underlying data, the question arises as to what situations the *LR* is likely to be highly sensitive to data. This is most likely to occur when the same data is used for multiple probability assignments in the same evaluation. Recall that we are interested in the sensitivity of the *LR* to the underlying data and not to a specific factor. Imagine that the scenario given at the beginning of the methods section changed to:

> *A suspect (Mr. S) is accused of digitally sexually assaulting the victim (V) at some point during a party. Swabs of Mr S's fingernails are taken and reveal a DNA profile that has the same characteristics as Ms V. Mr S denies the digital assault and denies being at the party. He concedes that he was at a different party with his friend later that night and that his friend had previously attended the same party as the victim. He suggests that the victim's DNA may have transferred to his friend's hands from social interaction and then to his hands from social interaction. The source of the DNA is thus not in dispute.*

The propositions do not change, but the updated BN structure is shown in Figure 6.



5. V DNA transfer to F fingernails from social interaction		Yes	No
6. V DNA transfer from F fingernails to S fingernails from social interaction	Yes	0.18	0
	No	0.82	1

425 *Figure 6: BN constructed for evaluation of DNA results given updated sexual assault scenario*
426 *explained in the main text (left) with additional node probability table (right). In the probability*
427 *table the node to which the individual table applies has been bolded.*

428

429 In the probability assignments for the BN is Figure 5 the Goray et al. [34] data is used for both
430 nodes 5 and 6. A rederivation of the *LR* formula for this new scenario is:

431 $$LR = \frac{1 + \Pr\left(\overline{T_D}\right)\left[\Pr\left(T_S\right)^2 - 1\right]}{\Pr\left(T_S\right)^2}$$

432 and in fact, the scenario could be extended to any number of intermediatory social gatherings
433 sitting between the party at which the victim was allegedly assaulted and the party the suspect
434 attended by replacing the power of 2 with a power of X (where X is the number of parties). If we
435 now considered a situation where the results of the Goray et al. [34] study were that transfer in a
436 social setting was not observed (i.e., 0 out of 48 observations) then the $LR \approx 1{,}750$ and the lower
437 bound is 133. This is a factor of 13 difference and so above an order of magnitude.

438

4.0 DISCUSSION:

In a practicing forensic laboratory, that must meet report deadlines and organisational budgets, it will most often be the case that bespoke experiments will not be an option to address a highly sensitive *LR*. Some testing of the robustness of the evaluation is an important step to take in order to diagnose any potential for insufficient knowledge unduly overstating the value of the evidence.

4.1 – Why a definition of highly sensitive is needed

In section 3 we stated that we believe it is a requirement within a practicing forensic laboratory to develop some definition for categorising when the *LR* is highly sensitive to a particular probability assignment (based on the underlying data sample size). An opinion that is sometimes expressed is that there need not be a definition of when an *LR* is highly sensitive, and instead the sensitivity of the *LR* to the underlying data can be discussed or explained to stakeholders. We agree with this sentiment in theory but find in practice it is problematic. The first major hurdle is to explain the concept of sensitivity to a stakeholder and convey in lay terms the effect of the data on the *LR* assignment. Even if this task were perfectly achieved, the obvious question coming from the discussion (or an obvious part of the explanation) would be '*is that sensitivity good or bad?*'. In other words, questioning whether the sensitivity analysis suggests the result is robust. This then naturally leads to the need for some criteria by which sensitivity analyses are deemed to show that the evaluation is robust, or not robust.

This idea is expressed in other works, such as in Taylor et al. [16]:

"*Thus, exploring the sensitivity of the LR to the underlying data allows us to explore whether our knowledge is sufficient to ensure robust conclusions.*"

which implies there must be some criteria by which to judge that a conclusion is not robust. More explicitly from the same paper:

"*…guidelines will need to be developed within the laboratory, based on the properties of the sensitivity analysis, in order to define what type of results can be reported.*".

In sub-source evaluations of DNA profile data, the issue of defining a 'sensitivity threshold' is negated by carrying out a sensitivity analysis (for example to the data informing allele frequencies,

or Fst [19, 26], or to the probability assigned to dropout [35]) and reporting the least discriminating *LR*. We deal with reporting options in the next section.

4.2 – How to report a highly sensitive *LR*

There are several potential options for what to do when an *LR* is deemed highly sensitive (by whatever means this is defined):

- Report the *LR* corresponding to the least discriminating bound,
- Report the mean *LR* only in the standard manner,
- Report the mean *LR* and provide a caveat on its sensitivity or provide its range,
- Report that a robust *LR* cannot be provided.

Whilst the choices that surround the designation of an evaluation as highly sensitive or robust is to some extent arbitrary, such decisions need to be made in order for a laboratory to function. Certain practices can be implemented that alleviate *LR* sensitivity to data, such as the use of a hyperprior count vector of (1, …, 1) for all counts. This does not address probability assignments that are based on modelling distributions, but this level of complexity being used in evaluations in practice is rare, and beyond the scope of this paper. Given these ameliorating practices it is likely that the only time sensitivity will be an issue is when the same data has been used to assign multiple probabilities.

One path forward is to calculate the 95% bounds on the sensitivity analysis for each factor (node or probability) in the evaluation. If the lower bound on any of the sensitivity analyses is more than the laboratory's chosen threshold below the LR obtained using the case-assigned probabilities, then the evaluation may be considered highly sensitive. If the laboratory has the facilities to do so then another option is to carry out a single sensitivity analysis on all data used in the BN at once, and make a determination on robustness based on these results alone. Doing a single, all-encompassing sensitivity analysis is likely to show the greatest possible level of sensitivity in the LR as it subsumes the sensitivities to each individual dataset. Again, based on this result it may be the case that the evaluation is considered highly sensitive. In such an instance, one of the options listed above is to decline to provide an evaluation (or to report that the findings cannot be meaningfully assessed). This choice has some problems. First, doing this simply shifts the burden of the choice of what activity occurred to the Court, unaided by scientific evaluation. Second, even

when an evaluation is deemed highly sensitive, there may still be discrimination power in the analysis. For example, imagine that for an evaluation the *LR* was deemed to be 50 000 and the sensitivity analysis showed a range for the *LR* that lay between 5 000 and 500 000 in support of one of the propositions, then even at its lowest the evaluation still has substantial power to aid the fact finder.

One option is to report the least discriminating bound (in the example, 5 000) without further mention of sensitivity analysis. This is akin to the commonly used reporting methodology for sub-source DNA evaluations. This can be accompanied by a note in the report that the evaluation is sensitive to the underlying data and that measures have been taken so that the strength of evidence is not overstated.

Another option would be to report the *LR* but provide some information to the Court that the analysis is sensitive to underlying data. One option could be:

> *The probability of the findings in this case are 50 000 times higher if the events occurred in the manner described by prosecution rather than the manner described by defence. In order to assign the LR I have relied on scientific data and personal knowledge, both of which can be an imperfect description of reality. With every evaluation, I also carry out an assessment of the impact of the potentially imperfect knowledge on the size of the LR, called a sensitivity analysis. In this case the sensitivity analysis indicated that the evaluation was sensitive to the underlying data for assigning a probability to DNA transfer occurring during a handshake. In simple terms, had another similar study been available or set of knowledge been relied on then my sensitivity analysis indicates the LR could have reasonably been as low as 5 000, or as high as 500 000. However, the value for the LR that best represents my current knowledge is 50 000.*

The previous example is quite different to the case where the evaluation led to an *LR* being assigned a value of 5, and the sensitivity analysis demonstrated a 95% range between 0.001 and 500. If the evaluation is deemed to be highly sensitive and the lower bound crosses $LR = 1$ then we suggest reporting that the evaluation is sensitive to the data and that an informative evaluation cannot be provided. Importantly it should then be stressed that this result should be taken as the

observations providing no support for either proposition (i.e., is neutral), rather than being considered as the evaluation being absent. A possible reporting option could be:

> *In order to assign an LR I must rely on scientific data and personal knowledge, both of which can be an imperfect description of reality. With every evaluation I also carry out an assessment of the impact of the potentially imperfect knowledge on the size of the LR, called a sensitivity analysis. In this case the sensitivity analysis indicated that the evaluation was sensitive to the underlying data for assigning a probability to DNA transfer occurring during a handshake. In simple terms, had another similar study been available or set of knowledge been relied on then my sensitivity analysis indicates the LR could reasonably support either the events occurring in the manner described by prosecution, or the manner described by defence. Given this sensitivity I consider the findings neutral with regards to the propositions i.e., they do not provide support for either proposition.*

4.3 – Assistance carrying out sensitivity analysis

Many of the methods we have described in this paper go beyond the use of excel and BNs. This limits their ability to be used to those individuals who are able to write computer code. To help with this situation a website that hosts a R Shiny app has been created that accepts Bayesian network .net files and provides an automated resampling method of sensitivity analysis (as described in section 3.2), to produce a result similar to that seen in Figure 4. It also allows to conduct sensitivity analyses with binary state nodes (as described in section 3.1). The application runs directly from http://cchampod.shinyapps.io/BN_sensitivity or can be installed locally. The application code is on doi.org/10.5281/zenodo.7844553

4.4 – The importance of understanding limitations and assumptions in an evaluation

In this work we defined a relatively narrow scope, particularly for the paired data situation, where we assume the data from experimentation perfectly align with the case circumstances and the only uncertainty came from the size of the dataset used to inform a probability assignment. In practice

there will always be multiple unknowns that must be considered. These unknowns can be grouped into

- data applicability unknowns,
- data sampling variation and
- case unknowns.

The data applicability unknowns relate to the applicability of the data to the case they are being applied. For example, even if we knew the exact position that the suspect placed his hands on the victim during a social interaction it may be that the study does not specify an exact type of social interaction. In this instance it is difficult to know how applicable the study data is to the case, or if an adjustment is needed, in which direction to adjust. It may be that the study is explicit about how the social interactions were experimentally carried out, but the number of experiments carried out was limited, and had the same study been repeated (in exactly the same way) a different set of observations would have been obtained. This latter issue is data sampling variation and has been the main focus of our paper. Lastly, it may be that the study is quite clear as to how it was conducted, and what occurred, but the case circumstances are unknown. For example, even in our simple case scenario we could consider that there are many unknown aspects of the social interaction between the victim and suspect; how they interacted, how close they were and for what length of time, whether they were perspiring, and many others.

Faced with all these challenges it may at first seem that activity level evaluations cannot be performed. However, there are several points to remember. First, probability is the language of uncertainty and methods such as the use of Bayesian reasoning are designed to express our uncertainty about the world. Sensitivity analyses take this even further by showing us what aspects of an evaluation are particularly critical to the robustness of our opinion. We have provided several methods for doing so here, and the variants of these may be applied in different situations. Second, it is important to consider the role of a forensic scientist carrying out an activity level evaluation. The purpose of an activity level evaluation is to take the DNA profiling results, use the best available forensic data and try to place the results in a case context for the court. There are always difficulties in doing this. However, even with incomplete knowledge and limited data a forensic scientist is always in a better position than a lay person to advise on the significance of the observations. To take the stance that there are too many factors to consider would deprive the courts of much-needed expert advice. The author's personal experience in court is that they would much rather have the scientist's expert thoughts on a topic (with the limitations explained) than be

given nothing. That being said, it is also the role of the scientist to resist entreaties (real or perceived) to provide thoughts that are not expert.

The important point is that the scientist should not overstate the strength of the evidence, they should not hide the source of opinions or obscure the limitations in the evaluation. We have not gone into detail about the reporting of an evaluation (other than some examples of concluding statements), however we stress the importance of making clear the source of data used to inform probability assignments, and their limitations in a report. This is particularly important to highlight when the framework of circumstances surrounding the alleged crime are uncertain, when the study from which data has been obtained does not completely pair with the case circumstances, or if the scientist has had to use their personal judgement heavily in a probability assignment due to a lack of available data to inform them. For this last situation in particular it is important to highlight the limitations of the evaluation, so as not to present a result with an appearance of scientific precision (or authority) that is not deserved. It should be made clear that it is a subjective opinion, based on personal experience and the reasoning for the probability assignment given should be made transparent.

In forensic science there is often a pursuit of what people refer to as objectivity. We note that as long as humans are involved in carrying any evaluation or judgement there is always subjectivity, even if all people ultimately provide the same opinion. We believe that underlying the call for objectivity is really a call for accuracy and precision (or a reduction in the amount of 'noise' in an evaluation). As said by psychologist Daniel Kahneman "*As long as it's judgement, you can reduce noise but you cannot eliminate it*" [36]. One method that Kahneman suggest for reducing noise are to standardise decision-making through the use of algorithms, checklists or standard procedures. Another is to ensure individuals use similar methods of data location, interpretation and use. Consider how these apply to activity level evaluations. By being open to the methods used to locate and use data (or knowledge, or experience) then the mechanics of the evaluation is open to judgement and adoption by others. By setting up standard methods for carrying out sensitivity analyses and interpreting the results, there will be a consistency between analysts when providing opinion on robustness. Both of these measures reduce 'noise' in statements given about evidential meaning, particularly compared to the alternative (when no normal evaluation has been carried out) of providing ad-hoc and unprepared answers given in response to questions posed whilst on the stand (see Vuille et al [37] for a discussion on this opinion).

We have provided two methods for dealing with a situation where a lack of information leads to range of data sources being sought, which provide a range of probabilities that could be assigned. One method integrates the evaluation over a prior distribution for the probability. The other assigned a specific value but tests the sensitivity of the LR to the assignment across the plausible range. Our experience is that some individuals prefer the first course of action as they feel it to be the best (and some claim only) way to deal with the uncertainty. Others prefer not to use the first method as it leads to an *LR* value and gives (in their opinion) an illusion of precision to the *LR* and (again, in their opinion) hides the uncertainty. We feel that both methods have merits and situations may arise where one feels more appropriate than the other. Again, the important point is transparent communication about what has been done.

For more guidance on how to address these limitations and report them we direct the reader to [20].

5.0 CONCLUSION:

We have provided examples of sensitivity analyses using a simple example scenario. The same processes applied to single factors (or nodes) shown here can be applied to multiple factors (or nodes) in a larger evaluation without increased theoretical complexity of the task (they just become more numerous to create and interpret).

Conceptually, the simplest situation arises when the analyst has data which are perfectly aligned (or paired) with the case circumstances. In this situation the uncertainty to be dealt with is only that of the sample size of the study that produced the data. If this is the case, then there are steps that can be taken that ameliorate the issue of evaluations being highly sensitive to underlying data.

If the scientist relies on their knowledge (or multiple secondary data sources) to define a reasonable range for the probability assignment a sensitivity analysis can still be undertaken by varying the probability across the range and noting the effect on the *LR* (as shown in Figure 4). However, in this situation the need for a sensitivity analysis may be negated if the uncertainty in probability assignment is directly expressed as a distribution in the network itself (as shown in Figure 5).

If a sensitivity analysis is carried out, then three questions may arise:

1. If there are no clear bounds by which to limit the probability being examined, how can a sensible range be defined?
2. Once some sensible bounds on the probability have been defined, and the sensitivity in the *LR* investigated, how do we determine whether the evaluation is robust?
3. Once we determine whether an evaluation is robust, how should the result be reported?

Regarding the first of these we showed in section 3.1 how a standard 95% bound can be applied to a factor with binary states, given a number of data points from an underlying study. If the factor is not binary then a resampling scheme is required, but again 95% bounds can be used. Both of these methods have been implemented in the Shiny app associated with this paper.

The second point is likely to cause the most controversy amongst Bayesian purists, who would quite rightly point out that theory tells us there is no need to categorise an evaluation as sensitive or insensitive, as the uncertainty is inherent in the probabilities that are assigned and the LR thus expresses the uncertainty. While we agree with this view from a theoretical perspective, in practise we have found that the needs of stakeholders (with regards to understanding the evaluation they have been provided), and the needs of the forensic institution carrying out the evaluation (with regards to having standardised evaluation and reporting guidelines that maximise the available resources) necessitate some definition of high sensitivity. The simplest method, and the one which we have used in this work, is to define 'highly sensitive' as being when the ratio of the mean *LR* and the lower bound *LR* differ by more than some pre-defined amount.

With regards to reporting, we suggest that if the results of the sensitivity analysis find that the level of sensitivity is within the acceptable level defined by the laboratory then there need not be any change in the manner in which the evaluation is reported. If a result is deemed to be highly sensitive then there are several approaches that could be taken that include; report the *LR* corresponding to the least discriminating bound, reporting the mean *LR* only in the standard manner, reporting the mean *LR* and provide a caveat on its sensitivity or provide its range, or reporting that a robust *LR* cannot be provided. We provided some examples of wording that may be used in section 4.2.

We have discussed the sensitivity of the *LR* to the underlying data. A further aspect to this is the sensitivity of the LR to the conditioning information. An example could be the sensitivity of the *LR* to any assumptions made on unknown or uncertain aspects of the case circumstances [28]. As changes to the conditioning information effectively require a new evaluation, this falls outside of

the scope of sensitivity analyses as discussed in this paper. We suggest that any uncertain or unknown (but relevant) aspects of the case circumstances are discussed with the mandating authority. The outcome of such a discussion may be that the impact of certain assumptions on the outcome of the evaluation is to be made transparent to the users of the report. In such instances the report may include multiple evaluations under different sets of conditioning information.

As well having provided some practical guidance to assist forensic institutions with carrying out, interpreting and reporting the results of sensitivity analyses we also provide a tool that has been developed to assist with carrying out the resampling method of sensitivity analysis. This tool takes the commonly used .net file that describes a BN and allows the user to explore the effect on the *LR* of the underlying datasets. In combination this app and paper we hope will provide forensic institutions with the information and tools to interrogate the sensitivity of their activity level evaluations and thereby provide more robust and informed opinions.

REFERENCES:

[1] Gill P, Brenner CH, Buckleton JS, Carracedo A, Krawczak M, Mayr WR, et al. DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. . Forensic Science International. 2006;160:90-101.

[2] Gill P, Gusmão L, Haned H, Mayr WR, Morling N, Parson W, et al. DNA commission of the International Society of Forensic Genetics: Recommendations on the evaluation of STR typing results that may include drop-out and/or drop-in using probabilistic methods. Forensic Science International: Genetics. 2012;6:679-88.

[3] Gill P, Hicks T, Butler JM, Connolly E, Gusmão L, Kokshoorn B, et al. DNA Commission of the International Society for Forensic Genetics: Assessing the value of forensic biological evidence - guidelines highlighting the importance of propositions. Part I: Evaluation of DNA profiling comparisons given (sub)source propositions. Forensic Science International: Genetics. 2018;36:189-202.

[4] Willis SM, McKenna L, McDermott S, O'Donell G, Barrett A, Rasmusson B, et al. ENFSI Guideline for Evaluative Reporting in Forensic Science. European Network of Forensic Science Institutes (available at http://enfsi.eu/sites/default/files/documents/external_publications/m1_guideline.pdf); 2015.

[5] Catoggio D, Bunford J, Taylor D, Wevers G, Ballantyne K, Morgan R. An introductory guide to evaluative reporting in forensic science. Australian Journal of Forensic Sciences. 2019;51:S247-S51.

[6] Buckleton J, Taylor D, Bright J-A, Hicks T, Curran J. When evaluating DNA evidence within a likelihood ratio framework, should the propositions be exhaustive? Forensic Science International: Genetics. 2021;50:102406.

[7] Cook R, Evett IW, Jackson G, Jones PJ, Lambert JA. A hierarchy of propositions: Deciding which level to address in casework. Science & Justice. 1998;38:231-40.

[8] Taylor DA, Bright J-A, Buckleton JS. The 'factor of two' issue in mixed DNA profiles. Journal of Theoretical Biology. 2014;363:300-6.

[9] Morrison G. Special issue on measuring and reporting the precision of forensic likelihood ratios: Introduction to the debate. Science & Justice. 2016;5:371-3.

[10] Morrison G, Enzinger E. What should a forensic practitioner's likelihood ratio be? Science & Justice. 2016;5:374-9.

[11] Curran J. Admitting to uncertainty in the LR. Science & Justice. 2016;5:380-2.

[12] Ommen D, Saunders C, Neumann C. An argument against presenting interval quantifications as a surrogate for the value of evidence. Science & Justice. 2016;5:383-7.

[13] Berger C, Slooten K. The LR does not exist. Science & Justice. 2016;5:388-91.

[14] Biedermann A, Bozza S, Taroni F, Aitken C. Reframing the debate: A question of probability, not of likelihood ratio. Science & Justice. 2016;5:392-6.

[15] Hout Avd, Alberink I. Posterior distribution for likelihood ratios in forensic science. Science & Justice. 2016;5:397-401.

[16] Taylor D, Hicks T, Champod C. Using sensitivity analyses in Bayesian networks to highlight the impact of data paucity and direct future analyses: a contribution to the debate on measuring and reporting the precision of likelihood ratios. Science & Justice. 2016;56:402-10.

[17] Nordgaard A. Comment on 'Dissmissal of the illusion of uncertainty on the assessment of a likelihood ratio' by Taroni F., Bozza S., Biedermann A. and Aitken C. Law, Probability and Risk. 2016;15:17-22.

[18] Taroni F, Bozza S, Biedermann A, Aitken C. Dismissal of the illusion of uncertainty in the assessment of a likelihood ratio. Law, Probability and Risk. 2016;15:1-16.

[19] Taylor D, Balding D. How can courts take into account the uncertainty in a likelihood ratio? Forensic Science International: Genetics. 2020;48:102361.

[20] Taylor D, Kokshoorn B. Forensic DNA Trace Evidence Interpretation: Activity Level Propositions and Likelihood Ratios (1st ed.): CRC Press; 2023.

[21] Taylor D. The evaluation of exclusionary DNA results: a discussion of issues in R v. Drummond. Law, Probability and Risk. 2016;15:175-97.

[22] Samie L, Champod C, Taylor D, Taroni F. The use of Bayesian Networks and simulation methods to identify the variables impacting the value of evidence assessed under activity level propositions in stabbing cases. Forensic Science International: Genetics. 2020;48:102334.

[23] Bright J-A, Lee S-I, Buckleton J, Taylor D. Revisiting the STRmix(TM) likelihood ratio probability interval coverage considering multiple factors. bioRxiv. 2021:2021.06.25.449960.

[24] Bright J-A, Taylor D, McGovern C, Cooper S, Russell L, Abarno D, et al. Developmental validation of STRmix™, expert software for the interpretation of forensic DNA profiles. Forensic Science International: Genetics. 2016;23:226-39.

[25] Buckleton J, Bright J-A, Taylor D. Forensic DNA Evidence Interpretation, second edition. Boca Raton, Florida: CRC Press; 2016.

[26] Taylor D, Bright JA, Buckleton J, Curran J. An illustration of the effect of various sources of uncertainty on DNA likelihood ratio calculations. Forensic Science International: Genetics. 2014;11:56-63.

[27] Samie L, Champod C, Delémont S, Basset P, Hicks T, Castella V. Use of Bayesian Networks for the investigation of the nature of biological material in casework. Forensic Science International. 2022;331:111174.

[28] Taylor D, Kokshoorn B, Hicks T. Structuring cases into propositions, assumptions, and undisputed case information. Forensic Science International: Genetics. 2020;44:102199.

[29] Taylor D, Biedermann A, Hicks T, Champod C. A template for constructing Bayesian networks in forensic biology cases when considering activity level propositions. Forensic Science International: Genetics. 2018;33:136-46.

[30] Plummer M. Bayesian graphical models using MCMC. RJAGS. 2012.

[31] Højsgaard S. Graphical Independence Networks with the gRain Package for R. Journal of Statistical Software. 2012;46:1-26.

[32] Scutari M. Learning Bayesian Networks with the bnlearn R Package. Journal of Statistical Software. 2010;35:1-22.

[33] Flanagan N, McAlister C. The transfer and persistence of DNA under the fingernails following digital penetration of the vagina. Forensic Science International: Genetics. 2011;5:479-83.

[34] Goray M, Oorschot RAHv. The complexities of DNA transfer during a social setting. Legal Medicine. 2015;17:82-91.

775 [35] Gill P, Haned H. A new methodological framework to interpret complex DNA profiles using
776 likelihood ratios. Forensic Science International: Genetics. 2013;7:251-63.
777 [36] Mance J. Noise: A Flaw in Human Judgment, DANIEL KAHNEMAN, OLIVIER SIBONY,
778 AND CASS. R. SUNSTEIN, London: William Collins, 2021, 464 pp., £25.00. Journal of Law and
779 Society. 2022;49:226-31.
780 [37] Vuille J, Luparia L, Taroni F. Scientific evidence and the right to a fair trial under Article 6
781 ECHR. Law, Probability and Risk. 2017;16:55-68.

782