Assessing evidence and testing appropriate hypotheses

Norman Fenton[1], Queen Mary University of London and Agena Ltd
3 October 2014

THIS IS A PRE-PUBLICATION DRAFT OF THE FOLLOWING CITATION

Fenton, N. E. (2014). Assessing evidence and testing appropriate hypotheses. *Sci Justice*. doi:http://dx.doi.org/10.1016/j.scijus.2014.10.007

Abstract

It is crucial to identify the most appropriate hypotheses if one is to apply probabilistic reasoning to evaluate and properly understand the impact of evidence. Subtle changes to the choice of a prosecution hypothesis can result in drastically different posterior probabilities to a defence hypothesis from the same evidence. To illustrate the problem we consider a real case in which probabilistic arguments assumed that the prosecution hypothesis "both babies were murdered" was the appropriate alternative to the defence hypothesis "both babies died of Sudden Infant Death Syndrome (SIDS)". Since it would have been sufficient for the prosecution to establish just one murder, a more appropriate alternative hypothesis was "at least one baby was murdered". Based on the same assumptions used by one of the probability experts who examined the case, the prior odds in favour of the defence hypothesis over the double murder hypothesis are 30 to 1. However, the prior odds in favour of the defence hypothesis over the alternative 'at least one murder' hypothesis are only 5 to 2. Assuming that the medical and other evidence has a likelihood ratio of 5 in favour of the prosecution hypothesis results in very different conclusions about the posterior probability of the defence hypothesis.

1. Introduction

Much has been written about the improper use of statistics in respect of Sudden Infant Death Syndrome (SIDS) in the Sally Clark case [1][2][5][6][7] (and others similar [8][9]). In particular, the claim made by the prosecution's key expert witness Roy Meadow at the original trial – that there was "only a 1 in 73 million chance of both children being SIDS victims" – has been thoroughly, and rightly, discredited. However, this was not the only statistical error made. Instead, there was failure to compare the prior probability of SIDS with the (also small) probability of murder[2]. Probability experts used this lack of comparison as the focus for discrediting the original statistical claims [2][6]. However, another statistical issue was not considered and this paper will present this issue using the Sally Clark case to illustrate an alternative view.

[1] n.fenton@qmul.ac.uk, Queen Mary University of London

[2] It is important to note that, although the focus of the statistical analyses have been on comparing the probabilities of SIDS and murder, these were not the only possible explanations for the deaths and this is now widely accepted. SIDS is partly a catch-all diagnosis for unexpected, unexplicable, spontaneous death of an infant while asleep. As stated in [6] the deaths could also have been the result of accidents or medical conditions that could have been explained with appropriate investigations. It was because no explanation was offered that SIDS was assumed to be ***the*** alternative explanation to murder. The implication of this assumption (which is now widely assumed to be untrue) is discussed further in the conclusions.

2. Errors in statistical arguments

2.1 Previously noted errors

There were two fundamental statistical/probabilistic errors which can be illustrated through the Sally Clark case:

1. The figure of 1 in 73 million for the probability of two SIDS deaths in the same family was based on an unrealistic assumption that two deaths would be independent events, and hence that the assumed probability of 1/8500 for a single SIDS death could be multiplied by 1/8500. The error is compounded when the resulting (very low) probability is assumed to be equivalent to the probability of Sally Clark's innocence (this is an example of the prosecutor's fallacy [3]).
2. The (prior) probability of a SIDS death was considered in isolation, i.e. without comparing it with the (prior) probability of the proposed alternative, namely of a child being murdered by a parent[3].

Although error 1 has been the most widely discussed, it is error 2 that makes the real difference to the case. Assuming, as in [6], that error 1 can be fixed by noting that a second SIDS death given a previous SIDS death in the same family is not 1/8500 but 5.7 times higher, i.e. 1/1491, it follows that the probability of two SIDS deaths is 1 in 12.6 million rather than 1 in 73 million[4]. To most lay people this difference might seem minor. That is why explaining and fixing error 2 was so critical, and this has been well documented in [1] (with longer version in [2]) and [6]. The key part of the explanation is based on comparing the prior probability of the SIDS hypothesis with the alternative murder hypothesis, i.e. specifically comparing the prior probabilities of the hypotheses :

M : Both babies were murdered.
S : Both babies died of SIDS.

In [1] the author provided the following argument:

> "In 1996 there were 649,489 live births in the England and Wales. On these babies, 14 were later classified as having been murdered in the first year of life. If we were to take the ratio 14/649,489 as our estimate of the probability that a single baby will be murdered in the first year of life, and manipulate it in exactly the same way as we did the SIDS rate, we would calculate that the probability of two babies in one family both being murdered is (14/649,489) times (14/649,489)[5], which gives 1 in 2,152,224,291. On this basis, the "logic" of paragraph 14 above would imply that we could essentially exclude the possibility that Sally Clark's two babies were murdered."

So, using the same independence assumptions that led to a "1 in 73 million" figure for the prior probability of *S*, we end up with a much lower prior probability of "1 in 2,152,224,291" for *M*. We are therefore able to conclude that the prior probability of *S* is actually 30 times greater than the prior probability of *M*.

[3] Although we have noted in footnote 1 that SIDS and murder should not have been the only alternatives, this paper focuses on previous statistical analyses of the case that did make this assumption.
[4] The data in [1] suggests the probability is higher - about 1 in 296400, while the probability of two natural deaths of infants in the same family is certainly much higher still.
[5] Note: In the paper there was a typo stating (4/649,489) times (4/649,489).

However, there is another hypothesis that could be considered. To explain it we will make the same basic simplifying assumption as in [1][2][6], namely that SIDS and murder are the only two possible causes of deaths in the case.

2.2 An alternative hypothesis and its probabilistic impact

The major problem illustrated above is in the formulation of the two hypotheses *M* and *S*, which were accepted as the only hypotheses. Although the author in [6] used exactly the same hypotheses *M* and *S* as in [1][2] he was clearly aware of the potential for error in doing so, because he also stated the following in the Introduction:

> "…there is the possibility that the deaths were a mixture of SIDS and homicides. With this proviso in mind, it still seems to me that the most relevant comparison to make here is that of the chances of multiple SIDS against the chances of multiple homicide, the reason being that when a case of multiple deaths comes to trial, it is generally the case that the prosecution claims that the deaths are all homicide, while the defence claims that they are all SIDS."

In fact, while the prosecution[6] might have *claimed* that both deaths were murder, it could have been sufficient for them to establish just one murder in order to get a convicton. This example highlights the importance of considering different hypotheses. In the Sally Clark case it was argued that only *M* should be considered, but a more relevant 'prosecution' hypothesis could be:

> *H*: ***At least*** one of the babies was murdered

Hence, it is hypothesis *H* and not hypothesis *M* that should be compared to *S*.

Note also that, unlike *M*, the hypothesis *H* is the logical negation of *S*; for reasons covered in detail in [4], the prosecution and defence hypotheses should ideally be logical negations of each other if we want to compare probabilities meaningfully when evidence is presented. What was considered in this case is only a partial hypothesis. Before we examine the statistical implications in the case, it is useful to illustrate its importance with a very simple hypothetical example:

> Suppose there are 100 raffle tickets numbered 1 to 100. Only tickets numbered 1 to 5 win a prize. Fred's alleged 'crime' is that he did not win a prize. The prior of the defence hypotheses H_d ('Fred selected a number less than 6') is a very low probability, namely 1/20, illustrating the strength of the prosecution case.
>
> Now suppose that we try to counter the low prior probability of H_d by arguing that we should compare it against the hypotheses H_p: 'Fred selected a number greater than 99'. The prior probability for H_p is 1/100, so this hypothesis is actually 5 times LESS likely than the defence hypothesis. While this is a valid conclusion for this particular hypothesis it is certainly not valid for the correct alternative hypothesis in this case, which is the logical negation of H_d :'Fred selected a number greater than 6' whose prior probability is, of course, 19/20. This is 19 times more likely than the defence hypothesis.

In Section 2 we already noted that [1] provided an estimate for the prior probability for hypothesis *M* as 1/2,152,224,291. In order to demonstrate the impact of using the

[6] Normally, it is the responsibility of the prosecution (only) to propose its hypothesis, such as 'defendant committed crime X'. If the prosecution cannot prove its hypothesis then the defendant is not convicted. So, by default, the defence hypothesis ('defendant did NOT commit crime X') is the logical negation of the prosecution hypothesis. The Sally Clark case was unusual in that there was an assumed defence hypothesis (*S*) and the prosecution (in the form of Meadow's expert testimony) challenged this hypothesis by arguing that its (prior) probability was very low. That is the reason why it makes sense in this context to focus on the defence hypothesis.

hypothesis *H* rather than *M*, we can use the same assumptions in [1] to get an estimate of the prior probability of *H* (using the same independence assumptions):

P(*H*) = P(at least one murder)
= P(double murder) + P(one murder, one SIDS)
= P(double murder) + P(child 1 murdered, child 2 SIDS)+
P(child 2 murdered, child 1 SIDS)
= 1/(2.15billion) + (1/8500 × 1/46,392) + (1/46,392 × 1/8500)
= 1/(2.15billion) + 1/394 million + 1/394 million
= approx 1/183 million

This has to be compared with the prior probability of 1/73 million for the defence hypothesis *S* using the same 'independence' assumptions. So, whereas the prior defence hypothesis *S* is ***30 times more likely*** than the hypothesis *M* it is ***only 2.5 times more likely than H***.

It is very important to point out that [1] was not supporting the idea that it made sense to assume independence for the SIDS deaths. The author was simply making the case that if we used the same (incorrect) independence assumption for the prosecution hypothesis that was used by Meadow for the defence hypothesis, then you ended up with an even lower probability. This is a valid and powerful way to get across the importance of error 2. However, as we have shown, the impact of error 2 is significantly exaggerated if one does not use the correct prosecution hypothesis: by allowing for the possibility of one murdered baby, the defense hypothesis weakens significantly.

2.3 The probabilistic impact upon additional evidence when the more appropriate hypothesis is used

It is beyond the scope of this short paper to consider the extent to which the medical and other evidence presented in the case changes the prior probabilities of the different hypotheses. What we can do, however, is apply the same hypothetical assumptions about the medical signs observed as done in [1] and compare the results when we start with the more appropriate hypotheses. The assumptions in [1] were:

> "..the probability that the specific medical signs observed would in fact be observed, on the hypothesis that the babies were murdered, was assessed at 1 in 20; while the probability of these same signs being observed on the hypothesis that they died of SIDS was taken to be 1 in 100. The ratio of these figures is 5."

Using the prosecution hypothesis *M* of a double murder only (so the defense hypothesis was 30 times more likely than that of the prosecution), it follows from Bayes' theorem[7] that the prior odds (of 30 to 1 in favour of the defence) swing by a factor of 5 in favour of the prosecution hypothesis (since the likelihood ratio is 5). This means that, even after taking acount of the evidence, the posterior odds are still strongly in favour of the defence hypothesis by 6 to 1. In contrast, using the alternative prosecution hypothesis *H* (where the defense hypothesis was only 2.5 times more likely) the swing by a factor of 5 in favour of the prosecution hypothesis results in posterior odds 2 to 1 ***in favour of the prosecution hypothesis***. So, with the same assumptions as were used in [1] there is a critical difference between using the hypotheses *M* (double murder) and the alternative hypothesis *H* (at least one murder). Whereas with the former the posterior probability of the defence hypothesis is still much more likely after observing the medical signs, with the latter the prosecution hypothesis becomes more likely. It is important to stress, of course, that the likelihood ratio of 5 (for the medical evidence) in favour of the prosecution hypothesis that was used in [1] was hypothetical; in reality, when all the evidence was considered the likelihood ratio may have been very different.

[7] In the odds form of Bayes' Theorem: *posterior odds = likelihood ratio × prior odds*

3. Summary and Conclusion

For a number of reasons presented in [4] when using probabilities to compare a pair of alternative hypotheses in legal cases, it is preferable that the pair are mutually exclusive and exhaustive. Indeed, in order to draw conclusions about the probative value of evidence directly from the likelihood ratio of the alternative hypotheses it is proved in [4] that the hypotheses ***must be*** mutually exclusive and exhaustive. This is also recognized as a requirement in [2]. This is why it is so crucial to identify the most appropriate hypotheses if one is to apply probabilistic reasoning to evaluate and properly understand the impact of evidence. Using the hypothesis "At least one baby was murdered" as opposed to "Both babies were murdered*"* makes a significant difference to the prior odds of the defence hypothesis ("Both babies died of SIDS").

It was beyond the scope of this paper to consider a serious statistical analysis of the Sally Clark case taking account of all relevant prior probabilities. This would involve access to extensive information about all types of child deaths, all known instances of multiple child deaths within the same family, and information about dependencies between different deaths.

4. Acknowledgements

I am indebted to the following colleagues for their valuable contributions and insights: Daniel Berger, Ray Hill, Anne Hsu, David Lagnado, Martin Neil, Pat Wiltshire. The work was supported in part by the European Research Council under grant ERC-2013-AdG339182-BAYES_KNOWLEDGE.

5. References

[1] Dawid, A. (2003). Sally Clark Appeal: Statement of Professor A. P. Dawid, http://www.statslab.cam.ac.uk/~apd/SallyClark_report.doc

[2] Dawid, P. (2005). Probability and Proof. Online Appendix to "Analysis of Evidence" T. Anderson, D. A. Schum and W. Twining. Cambridge University Press, http://tinyurl.com/tz85o

[3] Fenton, N. E. (2011). "Science and law: Improve statistics in court." Nature 479: 36-37.

[4] Fenton, N. E., D. Berger, D. Lagnado, M. Neil and A. Hsu, (2013). "When 'neutral' evidence still has probative value (with implications from the Barry George Case)", Science and Justice, http://dx.doi.org/10.1016/j.scijus.2013.07.002, Pre-publication draft: www.eecs.qmul.ac.uk/%7Enorman/papers/probative_value.pdf

[5] Forrest, A. R. (2003). "Sally Clark - a lesson for us all." Science & Justice: 43: 63-64.

[6] Hill, R. (2005). "Reflections on the cot death cases." Significance 2(13-15).

[7] Nobles, R. and D. Schiff (2005). "Misleading statistics within criminal trials: the Sally Clark case." Significance 2(1): 17-19.

[8] Meester, R., Collins, M., Gill, R., & van Lambalgen, M. (2007). On the (ab)use of statistics in the legal case against the nurse Lucia de B. Law, Probability and Risk, 5, 233–250.

[9] Sjerps, M., & Berger, C. (2012). How clear is transparent? Reporting expert reasoning in legal cases. Law, Probability and Risk, 11 (4), 317–329.