Sudden Infant Death or Murder? A Royal Confusion About Probabilities

Neven Sesardic

ABSTRACT

In this article I criticize the recommendations of some prominent statisticians about how to estimate and compare probabilities of the repeated sudden infant death and repeated murder. The issue has drawn considerable public attention in connection with several recent court cases in the UK. I try to show that when the three components of the Bayesian inference are carefully analyzed in this context, the advice of the statisticians turns out to be problematic in each of the steps.

1 Introduction

There has been a lot of publicity recently about several women in the United Kingdom who were convicted of killing their own children after each of these mothers had two or more of their infants die in succession and under suspicious circumstances. (A few of these convictions were later overturned on appeal.) The prosecutor's argument and a much discussed opinion of a crucial expert witness in all these court cases relied mainly on medical evidence but a probabilistic reasoning also played a (minor) role.

It was this latter, probabilistic aspect of the prosecutor's case that prompted a number of statisticians to issue a general warning about what they regarded

as the expert witness's serious statistical error, which they thought might have had a profound influence on the jury. Among those who intervened and tried to instruct the public and the courts about the correct use of probability were the Royal Statistical Society (RSS), the President of the Royal Statistical Society, the President of the Mathematical Association and several prominent academics, including two professors of statistics (from Oxford and University College, London). In my opinion, their advice about how to proceed with statistical reasoning is actually problematic in each of its three inferential steps.

For reasons that do not concern us, it was one of these court cases (*R v. Clark*) that caught most of the attention of the media. The solicitor Sally Clark's child, Christopher, died in her presence in November of 1996, 11 weeks after birth. The death was ascribed to respiratory infection. About a year later, her new baby, Harry, also died while he was alone with his mother and when he was 8 weeks old. At this point Clark was accused of murdering both infants and then in 1999 sentenced to life imprisonment. Her first appeal in 2000 was unsuccessful but on the second appeal in 2003 she was acquitted, principally because the judge decided that the surfacing of some previously undisclosed microbiological information that was never considered at trial has made the conviction 'unsafe.'

I have no intention here to argue about Sally Clark's guilt or innocence. My only goal is to evaluate the recommendations of some leading statisticians about how to make the appropriate probability judgment in this kind of situation. I am aware, of course, that many a reader will have little patience for listening to a little-known philosopher who announces that the Royal Statistical Society is wrong about—statistics! Well, I must confess to having a strange feeling myself about being engaged in such a quixotic enterprise, but I can only hope that this will not make you stop reading and throw this article away.

2 Setting the Stage: Bayes's Theorem

This section should be rather uncontroversial. It introduces a basic equation from the elementary probability theory (called 'Bayes's theorem'), which is relevant for discussing different views on Sally Clark's case. The question at issue has often been phrased as deciding between the following two alternative explanations of the deaths of the two children: (i) that both cases were instances of the sudden infant death syndrome (SIDS) or (ii) that the children were killed by their mother. This does not correspond to the way the debate was actually conducted in court, but since the statisticians I will critique framed the central issue in these terms (Dawid [2002]; Joyce [2002]; Hill [2004]), I will follow their presentation.

If we call these two hypotheses 2S (double SIDS) and 2M (double murder) then with E being the relevant empirical evidence presented in the court, the 'odds form' of Bayes's theorem gives the following relation between the posterior probabilities of the two rival hypotheses:

$$\frac{p(2S/E)}{p(2M/E)} = \frac{p(2S)}{p(2M)} \times \frac{p(E/2S)}{p(E/2M)} \tag{1}$$

Now according to the standard formula for calculating the probability of joint events, the probability of two cases of SIDS in one family is obtained by multiplying the probability of SIDS in general with the probability of a second SIDS, *given that one SIDS already happened.* That is:

$$p(2S) = p(S) \times p(S_2/S) \tag{2}$$

Obviously, the same logic applies to a double murder scenario, for which we can then write:

$$p(2M) = p(M) \times p(M_2/M) \tag{3}$$

Substituting Equations (2) and (3) into (1) yields:

$$\frac{p(2S/E)}{p(2M/E)} = \frac{p(S)}{p(M)} \times \frac{p(S_2/S)}{p(M_2/M)} \times \frac{p(E/2S)}{p(E/2M)} \tag{4}$$

Equation (4) is very useful because it shows that the ratio of the overall probabilities of the two competing hypotheses (2S and 2M) depends entirely on three ratios: (i) the ratio of prior probabilities of SIDS and murder, that is, $p(S)/p(M)$; (ii) the ratio of probabilities of repeated SIDS and repeated murder, that is, $p(S_2/S)/ p(M_2/M)$; and (iii) the ratio of the so-called 'likelihoods' of the two hypotheses, that is, $p(E/2S)/p(E/2M)$. The likelihood of a hypothesis is a term commonly used for the probability that a hypothesis confers upon the evidence (i.e., for how likely the evidence is on the assumption that the hypothesis is true).

How to estimate these three ratios? After carefully analyzing each of them separately I have concluded that a counsel about these matters coming from, or under the auspices of, the Royal Statistical Society is seriously misguided.

3 Prior Probabilities of Single SIDS and Single Homicide

It should be clear at the outset that 'prior' probabilities of SIDS and murder do not refer to probabilities of these events that are completely a priori. They are prior only in the sense that they represent probabilities *before* the concrete evidence pertaining to a particular case (say, Sally Clark) is taken into account. In other words, prior probabilities are also based on empirical evidence, but of a nonspecific kind (mostly data from population statistics).

Hence a very natural suggestion seems to be that the ratio of prior probabilities of SIDS and murder, p(S)/p(M), should be obtained in a standard way, by first taking the recorded incidence of SIDS and of infanticide in the relevant population, and then simply dividing the former magnitude by the latter. Indeed, several statisticians follow this procedure, apparently not being aware that in this particular context the empirically observed frequencies cannot be so straightforwardly translated into probabilities. The problem with this approach is that it leads to a wrong estimate of both probabilities, p(S) and p(M), and, to make things worse, the two mistakes go in opposite directions: p(S) is overestimated while p(M) is underestimated. As a result, the ratio of the two prior probabilities is seriously overstated.

Let us first see how Philip Dawid (Professor of Statistics at University College London) derived estimates for p(S) and p(M). His route to p(S) uses one of the most carefully conducted empirical studies of SIDS. According to the widely cited and praised Confidential Enquiry for Stillbirths and Deaths in Infancy (CESDI study), in families like Clark's (affluent, parents nonsmokers, mother over 26) the incidence of SIDS in the period 1993–1996 was 1 in 8,543 (Fleming *et al.* [2000], p. 92). Dawid takes this as the prior probability of SIDS for the case at hand.

How about p(M)? Here Dawid relies on the data from the Office of National Statistics for 1997, and claims that in that year there were seven cases in England and Wales in which children were killed in their first year of life. He then says that, given the total of 642,093 live births that year, 'this yields an estimate of the probability of being murdered, for one child, of 1.1×10^{-5}' (Dawid [2002], p. 76). This is one chance in about 92,000. So, the ratio of the prior probabilities that we are looking for, p(S)/p(M), would be around 11. In other words, the conclusion is that before the specific evidence in the case is taken into account, the probability of one of Clark's children dying of SIDS was eleven times higher than the probability of being murdered.

I found Dawid's figure of seven infant homicides in 1997 questionable because other scholars (e.g., Marks and Kumar [1993], p. 329; Fleming *et al.* [2000], p. 128; Levene and Bacon [2004], p. 443) say that the number of such cases per year in England and Wales has been consistently around 30. Therefore I checked Dawid's source, the report on mortality statistics for 1997 (Office of National Statistics [1998], pp. 174–7), and the mystery was easily resolved.

In the categorization of the Office of National Statistics (ONS), there are two codes for causes of death that are relevant here. First, E960–E969 refers to 'homicide and injury purposely inflicted on by other persons,' and second, E980–E989 refers to 'injury by other and unspecified means, undetermined whether accidentally or purposely inflicted.' In 1997, there were seven cases of infant deaths that were subsumed under the former code and 17 cases that were assigned the latter code. It therefore appears that Dawid arrived

at his estimate of the probability of infant homicide by including only the first group in his calculation and disregarding the second group completely. This is odd. The code E960–E969 covers the *established* homicides (i.e., those officially confirmed by the coroner or criminal court), and the code E980–E989 comprises *suspected* homicides (most of these cases are so classified pending the decision of the relevant authorities). Now it should be immediately clear even to a statistically unsophisticated person that at least some of the suspected homicides will eventually end up in the category of established homicides. Dawid's apparent omission of these 17 cases is bound to underestimate the true probability of infant homicide.

The effect of this 'misunderestimation' (pardon the Bushism) is far from negligible. Dawid's own source, the Office of National Statistics, actually recommends that in order to get the best estimate of the number of homicides, the suspected homicides should virtually all be treated as real homicides: 'A better estimate of homicides in years 1993 to 1998 may be obtained by *combining* ICD codes E960–E969, E980–E989 with inquest verdict ''pending''.' (Office of National Statistics [2005], italics added) The same advice comes from the CESDI study, which Dawid used for assessing the probability of SIDS: 'Deaths may be placed in the latter category [E980–E989] pending a coroner's verdict, *and most are eventually reassigned to the heading of homicide*. . ., (Fleming *et al.* [2000], p. 128, italics added).

So, this would mean that the number of infant homicides in 1997 should be increased from 7 to 24 (as the authors of the CESDI study also explicitly recommend). But even after this, a further correction is needed. ONS did not include homicides that happened in the first month of baby's life, so it follows that it still gives a too low figure for the infanticide frequency per year.[1] Besides, many scholars faced with all these uncertainties of estimation think that it is best to go directly to the official Home Office crime statistics, which lists the straightforward number of homicides per year as directly recorded by the police. According to this source (Home Office [2001], p. 87), there were 33 *recorded* killings of less than one-year old children in 1997, while the average number of these homicides per year in England and Wales in the period between 1990 and 2001 was 30. Since this figure is cited (among others) by the authors of the CESDI study, I also decided to rely on it in my further

---

[1] Speaking about the imperfect temporal match of our two databases (the CESDI source on SIDS and the ONS source on infant homicide), it should be mentioned that the CESDI study used the definition of SIDS that excludes infant deaths in the first *week* of life, but this 'shortening' of a 1-year period of data collection by 1/52 of its duration (less than 2%) is safely ignored. Introducing a correction for such a small discrepancy would just unnecessarily complicate our mathematics and yet the final result of our calculations would remain virtually the same.

calculations.[2] On this basis, Dawid's estimate for p(M) turns out to be wrong by a factor of 4.

Mathematician Ray Hill starts off on a better foot, citing the statistic of around 30 infant homicides a year among 650,000 births each year in England and Wales. But he also proceeds too quickly to translate the observed frequency into a probability statement: 'the chances of an infant being a homicide victim are about 1 in 21,700' (Hill [2004], p. 321). This estimate is also accepted by Helen Joyce in her paper ([2002]), in which she undertook to explain to the wider public how to use correctly the probabilistic reasoning in Clark's case. (Joyce was at the time the editor of *Significance*, a quarterly magazine published by the Royal Statistical Society.)

Are we ready now to obtain the ratio of prior probabilities, by dividing the probability of SIDS (Dawid's figure of 1 in 8,543) with the probability of infant homicide (Hill's figure of 1 in 21,700)? No, there are still two problems.

First, the observed frequencies from which the two probabilities are derived do not refer to the same group. Recall that Dawid's figure for SIDS refers to families like Clark's (affluent, parents nonsmokers, mother over 26). Hill's figure for infant homicide, however, refers to the whole population indiscriminately. It is well known that, like SIDS, infant homicide is also less frequent in the type of family defined by the three characteristics, but unfortunately I could not get a reliable statistic of infant homicide for that group. Therefore, in order to make the data for SIDS and infant homicide mutually comparable I decided to use the frequency of SIDS in the general population, instead of the more narrow category used by Dawid. According to the already mentioned study by Fleming *et al.*, the incidence of SIDS in the general population was 1 in 1,300.

Second, a serious methodological problem arises from the fact that SIDS is not a single well-defined syndrome, disease or type of event. It is a term of exclusion, or a 'dustbin diagnostic' (Emery [1989]). An essential component of the concept of SIDS is the ignorance of what caused the death. Simply, when the etiology of infant death is unknown then, if some additional conditions are satisfied, the death is classified as SIDS. In the most widely used formulation, SIDS is 'the sudden death of an infant under 1 year of age, which *remains unexplained* after a thorough case investigation, including performance of a complete autopsy, examination of the death scene, and review of the clinical history' (American Academy of Pediatrics [2000], p. 650, italics added).

---

[2]   Strangely enough, although Dawid has only words of praise for the authors of the CESDI study (which he calls 'a professionally executed study', Dawid [2002]), he does not even mention their advice to use the Home Office criminal statistics nor their decision to work with a more than three or four times higher figure for the incidence of infant homicide than his unreasonably low assessment.

This clearly opens up a possibility that a certain proportion of those cases that are labeled as SIDS are in fact homicides. Of course, the opposite possibility also exists, that some cases labeled as homicides are in fact SIDS, but for two reasons the latter kind of mistake is going to be much less frequent than the former. First, since a death will be classified as a homicide only if homicide is proved beyond reasonable doubt, miscategorizations will happen more rarely with cases labeled as homicide than with those labeled as SIDS, where inclusion is based largely on ignorance.[3] Second, a person who is typically the sole witness of a child death in those ambiguous 'SIDS or homicide' situations and whose report tends to be one of the most important pieces of evidence in the case will almost always try to prove the SIDS scenario. To put it bluntly, people who killed their children will argue that it was SIDS whenever there is minimal plausibility to that claim, but people whose children died of SIDS will seldom claim that they murdered them. This asymmetry will constitute another ground for taking the officially declared statistics of SIDS with some reservation and for not rushing into accepting it as an automatic reflection of the true probability of SIDS.

These are not just theoretical speculations. There are quite a number of cases that were classified as SIDS at first, but later ended in conviction for murder and even outright confession. Surely, there must also be cases of undiscovered murder that are wrongly treated as SIDS. How often does this happen?

It is hard to tell, but we can immediately reject the assumption implicitly made by Dawid, Hill and Joyce that the empirically collected data give an unbiased estimate of the actual incidence of SIDS and infanticide. They all take the recorded frequencies at face value, without taking into account the high probability that an unknown proportion of cases classified as SIDS are in fact covert homicides. To get a minimally reasonable assessment of the ratio of prior probabilities of S and M we would have to find a way to estimate the percentage of SIDS that are not true SIDS but something more sinister. This is not easy to do for the reasons explained above, but an additional difficulty is that in the case of infants even a very detailed autopsy cannot distinguish a SIDS death from a death caused by suffocation with a soft object.

Nevertheless, scholars have tried to get around the problem and obtain a reasonable ball-park figure for the proportion of homicides that are wrongly subsumed under SIDS.[4] John Emery suggested that perhaps between 10% and

---

[3] As John Emery said, SIDS is 'very easy to diagnose, as in general the less found the more certain the diagnosis' (Emery [1993], p. 1097).

[4] Notice that many actual but unsuspected homicides will be wrongly subsumed not only under the SIDS rubric but under *other* causes of death as well (accidents, various illnesses, etc.), and that this will lead to a further (and more massive) underestimation of the incidence of infant homicide: 'it is *generally acknowledged* that the known recorded cases of infant homicide are an underestimate of the actual number of infants killed' (Brookman and Nolan [2006],

20% of deaths attributed to SIDS may be unnatural deaths ([1993], p. 1099). The authors of the CESDI study considered the question and came up with the figure of 14% of SIDS cases in which maltreatment was thought to be either the main cause of death or a secondary or alternative cause of death (Fleming *et al.* [2000], pp. 126–7). In a recent article devoted to this very topic, it is suggested that, most plausibly, around 10% of SIDS are in reality homicides, while the proportion of covert homicides among deaths registered as 'unascertained' is probably even higher (Levene and Bacon [2004], p. 444). One of the latest comprehensive studies of homicide gives a figure of 20% as the best guess of the proportion of SIDS that are in reality nonnatural deaths (Brookman [2005], p. 21). Apparently, many pediatric pathologists and forensic pathologists say *in private conversations* that 'parental or adult intervention may have occurred in 20–40% of the cases of so called sudden infant death syndrome with which they are involved' (Green [1999], p. 697) but the publicly expressed opinions always give lower figures. Faced with this area of indeterminacy I decided to use estimates of p(S) and p(M) within a range of misclassified SIDS from 0% to 20%. The assumption of 0% error (all SIDS are true SIDS) is quite unrealistic, whereas many experts think that the assumption of 5–10% of misdiagnosis is unlikely to be a significant overestimate of the incidence of covert homicide.

The reader might think that a 10% error will only minimally affect our estimates of probabilities, and that it is hence somewhat pedantic to insist on introducing this kind of correction. Although this may be true about p(S) it is not at all true about p(M). Since there are so many more cases of reported SIDS than recorded infant homicides, even a small percentage of cases transferred from S to M will substantially increase the number of M cases. For instance, even if only 10% of SIDS are misdiagnosed, the correction will result in more than doubling of the probability of M. Look at Table 1 for details.

Let us run quickly through this table and explain its contents.

The first row (ΣS) gives the number of all SIDS cases (as recorded in the CESDI study) in its first column, and then in the subsequent columns this number is decreased by subtracting a percentage of false SIDS under each of the four different 'false SIDS scenarios' (5%, 10%, 15% and 20%).

The second row (ΣM) is based on the number of infant homicides per year from the Home Office crime statistics. Again, there is no correction in the first column because the assumption here is that there are zero cases of (covert) homicide wrongly subsumed under the SIDS rubric. In the subsequent

p. 870, italics added); '*Most* commentators claim that the official statistics probably *considerably underestimate* the true incidence of child homicide' (Wilczynski [1997], p. 25, cf. p. 36, italics added). I could not think of a sound way to assess the impact of this misclassification error. Therefore, the reader should keep in mind that my acceptance of a too low value for p(M) will have the direct effect of making my final estimate of the ratio p(S)/p(M) unduly inflated to an unknown extent.

columns, however, the number of homicides is increased by adding the corresponding number of incorrectly labeled SIDS that are in reality murders.

The third row (ΣM-corr) introduces another modification. Since the population which represents the reference group for the murder statistics (642,093 births in England and Wales in 1997) is considerably larger than the population for SIDS statistics (472,823 births covered in the CESDI study) it is necessary to add a correction factor to compensate for the difference in size between the two groups. Therefore, in each of the four cases of nonzero percentage of false SIDS, the number of estimated covert homicides that is added to the officially reported figure of recognized homicides has to be multiplied by 642,093/472,823, which is 1.36.

The fourth row gives the prior probability of S, by simply dividing ΣS (the first row) by 472,823.

The fifth row gives the prior probability of M, by dividing ΣM-corr (the third row) by 642,093.

The sixth row reveals the magnitude we are looking for in this section: the ratio of prior probabilities of S and M.

Several comments are in order here. If the reader is wondering why those two different sources were chosen for estimating probabilities (the CESDI study for sudden infant death and the government data for homicides) the answer is simple. It was not me who picked out these sources. They were used by the statisticians I critique. Recall that in this paper I am *not* trying to arrive at the most reasonable estimates of probabilities of S and M, all things considered, but rather to assess the statisticians' conclusions about these probabilities. My view is that their conclusions are badly off the mark *even when they are inferred from their own data.*

How wrong are the statisticians about p(S)/p(M)? Ray Hill overestimates the ratio by a factor of 2 or 3, if we accept the figure of 5 or 10% as the percentage of covert homicides passing as SIDS. Philip Dawid's error is more serious. Like Hill, he also takes at face value the statistical data about SIDS (neglecting the probability that an unknown proportion of case labeled as

Table 1 Frequencies and prior probabilities of S and M

	0%	5%	10%	15%	20%
ΣS	363	345	327	309	290
ΣM	30	48	66	84	103
ΣM-corr	30	55	79	104	129
p(S)	0.000768	0.000729	0.000691	0.000653	0.000614
p(M)	0.000047	0.000085	0.000124	0.000162	0.000200
p(S)/p(M)	16.4	8.6	5.6	4.0	3.1

SIDS are homicides). So when his additional miscalculation of incidence of infant homicides (1 in 92,000) in the general population is combined with the comparable estimate of p(S) as 1 in 1,300 in the general population, this yields the ratio of 70 for p(S)/p(M), an assessment that is wrong by a factor of 8 or 12 (again, on the assumption that the ratio of misclassified SIDS is 5 or 10%).

A clear result of our analysis is that p(S) is indeed significantly higher than p(M) under any of the percentage scenarios of misdiagnosed SIDS. In other words, the first component of the right-hand side of Equation (4) makes the probability ratio favorable for the double SIDS hypothesis.

Now after having obtained our first ratio, p(S)/p(M), we are ready to plug it into Equation (4), as the first step in determining the overall ratio of posterior probabilities of 2S and 2M. For concreteness, I will select the 10% assumption of misdiagnosed SIDS, and then for each stage of the discussion I will show how Equation (4) looks under that assumption. I chose this percentage because, as mentioned above, it comes from the most recent and most detailed analysis in the literature (Levene and Bacon [2004]). But remember that the figure of 10% is just an estimate, used as an illustration for those who have a preference for specific comparisons. For a more synoptic view, look at the graphs later in the text that contain more comprehensive information. (Also bear in mind the important warning spelled out in footnote 4).

With this caveat, here is how Equation (4) is modified after the first stage of our debate, with the ratio p(S)/p(M) replaced by the value 5.6 from Table 1 (the 10% column):

$$\frac{p(2S/E)}{p(2M/E)} = 5.6 \times \frac{p(S_2/S)}{p(M_2/M)} \times \frac{p(E/2S)}{p(E/2M)} \tag{5}$$

4 Prior Probabilities of the Recurrence of SIDS and Homicide

Now it is time to consider the second component of Equation (4), the ratio of prior probabilities of repeated SIDS and of repeated murder: $p(S_2/S)/p(M_2/M)$. It is in this context that the statisticians complained most about the treatment of probabilities in the Sally Clark court case. The focus was on Roy Meadow, an expert witness for the prosecution, who suggested that the probability of a double SIDS can be obtained by just squaring the probability of a single SIDS. In this way, starting with the odds of 1 in 8,543 for a *single* SIDS in the relevant type of family (affluent, parents nonsmokers, mother over 26), Meadow claimed that the probability of a *double* SIDS in that kind of family is 1 in 73 million ($8,543 \times 8,543$).

This squaring of the single SIDS probability caused an outrage among statisticians. First, the President of the Royal Statistical Society was approached by a Fellow who was involved in the campaign to clear Sally Clark and who asked 'that the Society consider issuing a statement on the

misuse of statistics in the court case' (Royal Statistical Society [2002a], p. 503). And indeed, the RSS decided to make a public statement about 'invalid probabilistic reasoning in court,' which referred to Meadow's handling of probabilities. What the Royal Statistical Society called 'a serious statistical error' was soon described as a 'howler' (Matthews [2005]), 'an elementary statistical mistake, of the kind that a first year student should not make' (Dalrymple [2003]), 'infamous statistic' (Hill [2005]), 'blunder' (Marshall [2005]), 'poor science' (the Oxford statistician Peter Donnelly, quoted in Sweeney [2000]), 'disgrace' (the geneticist Brian Lowry, quoted in Batt [2004], p. 244) and even 'the most infamous statistical statement ever made in a British courtroom' (Joyce [2002]).

Meadow was demonized in print to such an extent that, when he appeared before the disciplinary panel of the General Medical Council (GMC) in July of 2005, people might have been surprised to notice, as a journalist put it, that the 'disgraced' and 'discredited' 'child-snatcher in chief' walked on feet rather than cloven hooves (Gornall [2005]). The GMC promptly decided to strike Meadow off the medical register,[5] in the face of the isolated voice of opposition of the editor of *Lancet*, who argued that 'facts and fairness demand that Prof Roy Meadow be found not guilty of serious professional misconduct' (Horton [2005]). A court of appeals overturned the GMC verdict in February 2006, but the appeal of GMC to that decision was partly successful in October 2006.

Returning to probabilities, what exactly was the statistical sin that Meadow was guilty of? Here is the explanation of the Royal Statistical Society:

> This approach (the squaring of the single SIDS probability) is, in general, statistically invalid. It would only be valid if SIDS cases arose independently within families, an assumption that would need to be justified empirically. Not only was no such empirical justification provided in the case, but there are very strong *a priori* reasons for supposing that the assumption will be false. There may well be unknown genetic or environmental factors that predispose families to SIDS, so that a second case within the family becomes much more likely. (Royal Statistical Society [2001])

The same criticism is expressed in an open letter that the President of the Royal Statistical Society sent to the Lord Chancellor on January 23, 2002 (Royal Statistical Society [2002b]).

Now it is undeniable that in inferring the probability of two SIDS in the same family Roy Meadow proceeded as if $p(S_2/S)$ were equal to $p(S_2)$, that is,

---

5 In marked contrast, Solicitors Disciplinary Tribunal made an unprecedented decision in 2001 that Sally Clark should *not* be struck off the Roll of Solicitors, although at the time her conviction for infant homicide remained in place, as her second (successful) appeal had not yet been heard.

as if the probability of a second child dying of SIDS was not affected by the fact that the first child died of SIDS. Let us call this assumption that $p(S_2/S) = p(S_2)$, the independence hypothesis (IH). RSS is certainly right that Meadow's squaring of probability is valid only if IH is true. It seems to me that there are two possible explanations for why Meadow squared the probability: he either did it because (i) he did not realize that it is necessary to assume the truth of IH (thus making an elementary mistake in probability reasoning), or because (ii) he actually had reasons to assume that IH is true and relied on its truth in his inference. In case (ii) there is no statistical fallacy at all, and then the only question would be whether his empirical assumption is in fact correct or not.

It is very odd that neither RSS nor the President of RSS did as much as mention possibility (ii). They immediately opted for diagnosis (i), apparently for two reasons: first, because Meadow provided no empirical justification for IH, and second, because they thought there are strong grounds to believe that IH is in reality false.

But neither of these two reasons supports explanation (i). First, as an expert witness, Meadow was surely under no immediate obligation to provide specific empirical justification for his opinion about this issue, especially since he was not asked by anyone to elaborate, and since his view was not even challenged by the defense when he expressed it.

Second, when the statisticians assert that there are 'very strong *a priori* reasons' (RSS) or 'strong reasons' (the president of RSS) for supposing that IH is false, they really venture forth onto thin ice. They are right that 'there may well be unknown genetic or environmental factors that predispose families to SIDS, so that a second case within the family becomes much more likely.' But then again, there are also 'very strong *a priori* reasons' that pull in the opposite direction, making a second SIDS case within the family *less* likely. For instance, common sense tells us that the parents whose first child died of SIDS are likely to take extreme precautions in their care of the next child, be attentive to the smallest signs of child's discomfort, never or rarely leave it alone, change their own habits (e.g., quit smoking), strictly follow the doctor's advice about minimizing the risk of recurrence, etc., which would all have the effect of making the probability of a second SIDS in the same family *lower* than the probability of SIDS in the general population.

This shows that RSS was wrong when it claimed, on the basis of 'a priori strong reasons,' that the error resulting from accepting IH is likely to be 'in one particular direction' (i.e., that it will underestimate the probability of repeated SIDS). There is simply no way to know in advance which of the two kinds of a priori reasons pulling in opposite directions will carry more weight.

More importantly, however, since the truth-value of IH is an empirical issue, why should the discussion about it be conducted in nonempirical terms? Why start criticizing a very distinguished pediatrician's testimony about pediatrics

by disputing his claims with a priori reasons or by imputing to him an elementary mistake in the logic of probability? If the expert's expressed view in his own field of competence implicitly assumes the truth of IH, does not minimal fairness dictate that, before his commitment to IH is dismissed as being the result of sloppy and fallacious thinking, we first seriously consider the possibility that he actually possessed *good empirical reasons* to believe that IH is true? The statisticians who intervened in the Sally Clark case by attacking Meadow's testimony left this avenue completely unexplored.

An indication that Meadow's reliance on IH was not an ill-considered and precipitate judgment is that he defended it in his published writings as well: 'There do not seem to be specific genetic factors entailed [for SIDS] and so *there is no reason to expect recurrence within a family*' (Meadow [1997], p. 27, italics added). Is this a solitary opinion of a maverick expert, or is it shared by some other scholars in the field? Let us see.

The authors of a carefully designed 16-year study of the SIDS recurrence statistics in the state of Washington concluded: 'Parents of SIDS victims can be advised, with considerably more confidence than in the past, that their risk of loss of a child in the future is virtually the same as that among families of like size and mother's age' (Peterson *et al.* [1986], p. 914) The authors of another similar study conducted in Norway over 14 years and on a population of more than 800,000 children claimed that 'there is no significant increased risk [of SIDS] in subsequent siblings' (Irgens and Peterson [1988]). One of the leading SIDS researchers, J. Bruce Beckwith, writes: 'Until more satisfactory information is provided, I will continue to reassure families of these SIDS victims whose family history is negative, whose clinical and postmortem findings are classic for SIDS, and who are not burdened with an excess of risk factors that their recurrence risk is not increased significantly over that of the general population' (Beckwith [1990], p. 514). In a widely cited study on the subject it is suggested that 'the chance of recurrence [of SIDS] is very small and probably no greater than the general occurrence of such deaths' (Wolkind *et al.* [1993], p. 876). The editor of the main journal in the field of child medicine says resolutely: 'Some physicians still believe SIDS runs in families. It does not—murder does' (Lucey [1997]). It is the opinion of the American Academy of Pediatrics that 'the risk for SIDS in subsequent children is not likely increased' (American Academy of Pediatrics [2001], p. 439). A chapter on 'SIDS and Infanticide' in a recent authoritative collection of articles about SIDS contains the following statement: 'No evidence has ever been presented to demonstrate that families that have suffered one SIDS death are legitimately at a greater risk of a second, to say nothing of a third, fourth, or fifth. On the contrary: anecdotal evidence supports the suggestion that, as Linda Norton MD, a medical examiner specializing in child abuse, said: "SIDS does not run in families—murder does" (personal communication)'

(Firstman and Talan [2001], p. 296). In a textbook of forensic pathology that is called 'the most comprehensive, definitive, and practical medicolegal textbook of forensic pathology today' we can read: 'SIDS deaths appear to occur in families at random. There is no evidence of a genetic etiology. Siblings of SIDS victims have the same risk as the general population' (DiMaio and DiMaio [2001], p. 327). Thomas G. Keens, professor of pediatrics at the University of California who specializes in SIDS research, gave an overview of the relevant empirical studies and concluded: 'Thus, the SIDS recurrence risk for subsequent siblings of one previous SIDS victim is probably not increased over the SIDS risk for the general population' (Keens [2002], p. 21).

I am not implying that all scholars would agree with the opinion voiced in these quotations, but obviously many leading SIDS experts believe that IH is actually supported by empirical evidence and that it is true. This completely undermines the RSS's condemnation of Meadow's squaring of probabilities. Recall that it was on the basis of some a priori reasons (which *they* regarded as cogent) that the statisticians thought that IH is most probably false, and then on that assumption they concluded that the squaring of probabilities, which depended on the truth of IH, must have been the result of a crude mathematical error. But, as the above list of quotations proves, the statisticians are simply wrong in their judgment that it is so easy to dismiss IH on intuitive grounds. Ironically, the statisticians accused a pediatrician of venturing into a field outside of his area of expertise and committing a serious statistical fallacy, when what really happened was that the statisticians themselves left their own area of competence and issued a rather dogmatic and ill-informed pronouncement about a topic—in pediatrics[6]!

The authority of the Royal Statistical Society was sufficient to make this misconceived claim spread quickly in all directions and gain uncritical acceptance, sometimes in quite unexpected quarters. For instance, although many experts do argue that SIDS cases *are* independent (as documented above), the categorical but unsubstantiated statement that 'there is no scientific evidence' for the independence of SIDS deaths was even defended in the British Academy Annual Lecture that was given by the Governor of the Bank of England (King [2004], p. 10).

The intervention of the statisticians was entirely unnecessary because the legitimacy of squaring the probabilities in the Sally Clark case was never

---

[6]   In a talk in which he discussed the Sally Clark case the Oxford statistician Peter Donnelly also made a bare assertion that IH is 'palpably false' (Donnelly [2005]), without providing any actual empirical evidence against IH. Apparently he was unaware that he was thereby overstepping his authority as a statistician and that, moreover, his unsupported contention was resolutely rejected by many of the leading experts in SIDS research. Worse still, Donnelly failed to heed his own good advice from the same talk: 'We need to understand where our competence is and is not.'

really a mathematical or statistical issue. The multiplication stands or falls just depending on whether IH is true. And as we saw, many scholars in the relevant disciplines do accept IH, which shows that there was nothing atrociously wrong or irresponsible in Meadow's testimony invoking the odds of 1 in 73 million.

True, other experts reject IH, but even if IH is false, and if consequently $p(S_2/S)$ is higher than $p(S)$, the Royal Statistical Society is still wrong in its statement that in that case the figure of 1 in 73 million would involve an error that 'is likely to be *very large*' and that 'the true frequency of families with two cases of SIDS may be *very much less* incriminating than the figure presented to the jury at trial' (Royal Statistical Society [2001], italics added). The phrases 'very large' and 'very much less' seem to suggest that we have a good reason to expect a drastic increase of probability of SIDS, given a previous SIDS case. But the best studies in the literature do not support such dramatic jumps in SIDS risk at all. For instance, in a recent policy statement the American Academy of Pediatrics concluded that 'the risk of [SIDS] recurrence in siblings, if present, is most likely *exceedingly low*' (Blackmon *et al.* [2003], p. 215, italics added). Furthermore, to mention the most widely cited research reports, it has been suggested that the risk of SIDS in subsequent siblings of a SIDS victim is increased by a factor of 3.7 (Irgens *et al.* [1984]), 1.9 (Peterson *et al.* [1986]) or 4.8 (Gunderoth *et al.* [1990]). The most recent study on the topic gives the estimate of a 6-fold increase of the risk (Carpenter *et al.* [2005]), but as a perceptive critic pointed out (Bacon [2005]), this larger figure is probably an overestimate that resulted from two serious methodological flaws: first, the sample was not randomly selected (some families opted out from the investigation for unknown reasons), and second, some cases with a number of suspicious characteristics (possible homicides) were uncritically subsumed under the SIDS category.[7] The second objection actually also applies to the previously mentioned three studies, so that, all things considered, it could be argued that it is very unlikely that $p(S_2/S)$ is more than 4 times higher than $p(S)$.

Applying this estimate to the Sally Clark case, the prior probability of two infants dying of SIDS in that kind of family is obtained by multiplying 1/8,543 (the CESDI figure) with 1/2,136 (four times the CESDI figure), which gives the required probability as 1 in 18 million. Although this estimate (1 in 18 million) as opposed to odds of 1 in 73 million is clearly higher, it is still an astronomically low probability and it must be asked whether this would really

---

[7]   For example, the authors even classified as SIDS those situations where there were mental health concerns about parents, there was confirmed child abuse, and in 25% of all of these cases 'there were pathological findings compatible with an asphyxial component to the death.'

be regarded as 'very much less incriminating,' as the Royal Statistical Society suggested.

In our search for the ratio $p(S_2/S)/p(M_2/M)$ we are now in the position to weigh the number in the numerator: $p(S_2/S)$ is either the same as $p(S)$, or it is up to 4 times higher than $p(S)$. We will make use of both of these estimates later, at the final stage of the analysis in this section.

But first we have to try to get some idea about the magnitude of the denominator. Dawid has an odd proposal about how to proceed here. He says that since Meadow's obtaining the probability of double SIDS by squaring the single SIDS probability was based on the 'unrealistic' assumption of independence, then one might as well treat the probability of double murder 'in the same (admittedly dubious) way as the corresponding probability for SIDS, and square it to account for the two deaths' (Dawid [2002], p. 76; cf. Dawid [2004], p. 2). In this way he derives the conclusion that a double SIDS is much more likely than double murder.

Dawid's basic point is that the simple multiplication of probabilities is dubious in both cases, and that, therefore, if it is allowed in calculating the probability of double SIDS then it could, equivalently, be also applied in the case of double murder.

But in estimating the repetition probabilities in cases like Sally Clark's, the two scenarios that we are considering (double SIDS and double murder) are not equivalent at all. The crucial asymmetry consists in the fact that the first child's death was initially diagnosed as being due to natural causes, which means SIDS (in the perspective of 'either-SIDS-or-murder' that we adopted from the beginning). Now if it was really SIDS, the mere classification of death as SIDS could in no way increase the probability of the next sibling dying from SIDS. But if the first child's death was not a true SIDS but homicide, then the fact that it was *wrongly* classified as SIDS *would indeed substantially increase the probability of the next sibling being killed*. Why? Well, the wrong classification would have as a direct consequence that after the *undiscovered* homicide of the first child the next sibling would continue living together with the mother that already killed one baby, and under these circumstances the new child would be under grave risk.

So, squaring the single-case probability in order to get the recurrence probability is not equally problematic in the two cases (SIDS and murder). With SIDS, many scholars in the relevant area of empirical research think that the procedure is entirely legitimate. In the case of murder, however, most people would think that in the situation under discussion such a procedure would be manifestly wrong. It stands to reason that the probability of a child being killed *if it is in the care of a parent who already proved to be capable of killing a baby* must be significantly higher than the probability of a child being killed in general.

There will be exceptions, of course, in which an infant homicide may lead to, say, a Raskolnikovian guilt and to a character transformation of a murderer into a caring and devoted parent. Or a parent might be afraid that she would be caught if she did it again,[8] and she could refrain from killing again for that reason. This is all true but we are talking about probabilities here. Although the very logic of the situation makes it impossible to get access to the relevant statistical data (how could we know the recurrence rate of child homicide in families with a previous *undiscovered* infanticide?), indirect reasoning can help. Since it is well known that child abuse has a strong tendency to repeat itself, why expect that this suddenly stops with the ultimate abuse? Also, since it is well known that previous homicide offenders are much more likely to commit a homicide than people in the general population, why expect that this recidivist tendency abruptly disappears when victims of homicide are babies? Or to make the point with a rhetorical question: if the only thing you knew about persons A and B was that A intentionally killed her own child whereas B did not, which of the two would you prefer as a baby-sitter for your own son or daughter?

On this basis, we have to reject as misguided Dawid's idea (repeated and endorsed in Aitken and Taroni [2004], pp. 211–3) that if one squares the probability of a single SIDS to obtain the probability of a double SIDS, this somehow justifies doing the 'same' thing with the probability of murder. It is not the same thing. Inferring the probability of double S by squaring p(S) may well be regarded as controversial, but in the context that interests us, squaring p(M) to obtain the probability of double M is just blatantly wrong.

How should we quantify $p(M_2/M)$? We need some numerical value, however approximate, which we can plug in to obtain the ratio $p(S_2/S)/p(M_2/M)$. Mathematician Ray Hill, who was actively involved in the efforts to free Sally Clark, estimated that $p(M_2/M)$ is 0.0078 (Hill [2004], pp. 322–3). Since for reasons of space I cannot go into a detailed criticism of his derivation, let me briefly adumbrate four basic reasons why I do not find his figure credible. First, Hill used the data from a study that was based on voluntary participation, which raises serious concerns that the sample was not representative and that the biased selection led to the underestimate of the probability in question. Second, his calculation relied on the values of p(S) and p(M) that were directly estimated from observed frequencies, without correcting for a number of homicide cases that are expected to have been misdiagnosed as SIDS (the mistake that is extensively explained in Section 3). Third, Hill uncritically treated all cases classified as double SIDS as *really* being double SIDS, again disregarding an even more pronounced risk of wrong classification. Fourth,

---

[8] Notice, however, that there would not be much reason for such a fear in an environment in which multiple SIDS is regarded as a frequent and expected phenomenon.

there are reasons to think that the probability we are looking for will be considerably higher than Hill's figure of 0.0078. Namely, if a child lives with a parent who already killed another baby, it is arguable that the child's life is exposed to 'a clear and present danger' to such a degree that it is not adequately represented by the probability of its being killed that is as low as less than 1%.

Helen Joyce starts with an intuitive estimate of 0.1 for $p(M_2/M)$. Although she says that this figure 'is almost certainly overestimating the incidence of double murder' (Joyce [2002]), she nevertheless regards it as one of several 'reasonable estimates of the likelihoods of relevant events'. I will adopt Joyce's suggestion for $p(M_2/M)$ because I agree that it is not an unreasonable estimate, and because in this way I will follow my basic strategy of accepting those of the statisticians' premises that are not patently wrong and then assessing whether or not the conclusions they derive from them do actually follow.

Now we are ready to obtain the second component in Equation (4), the ratio $p(S_2/S)/p(M_2/M)$. On the assumption of statistical independence of SIDS cases in the same family, $p(S_2/S)$ is equal to $p(S)$. As before, for illustration purposes let us take $p(S)$ under the scenario of the 10% rate of misdiagnosed SIDS (covert homicide). In that case, $p(S)$ is 0.00069. If we divide this by Joyce's figure for $p(M_2/M)$, which is 0.1, we get 0.0069. This is the ratio of the repetition probabilities of SIDS and murder. If we plug this number into Equation (5), and multiply it with the earlier obtained ratio $p(S)/p(M)$, which was 5.6, the result is presented in the following Equation (6):

$$\frac{p(2S/E)}{p(2M/E)} = 5.6 \times 0.0069 \times \frac{p(E/2S)}{p(E/2M)} = 0.04 \times \frac{p(E/2S)}{p(E/2M)} \tag{6}$$

What is the meaning of 0.04 here? It represents the ratio of *prior* probabilities of the two hypotheses we are discussing (double SIDS and double murder), that is, *before* the specific empirical evidence in any specific case is considered. Put differently, the ratio tells us that a priori (before consulting the particular evidence pertaining to the case) the double murder hypothesis is 25 times more probable than the double SIDS hypothesis. Moreover, supposing that one of the two hypotheses must be true about a given unresolved child death, we can immediately obtain the prior probabilities of both hypotheses. The prior probability of 2M is 25/26, or approximately 0.96, whereas the prior probability of 2S is 1/26, or approximately 0.04.[9]

Of course, these particular probabilities apply only to a very specific situation, that is, under the assumptions that were made about $p(S)$, $p(M)$,

---

[9]   An attentive reader will ask: how come that if we divide $p(S)$ by $p(M)$ the result turns out to be 1/24 (0.04/0.96 = 1/24), rather than our previously calculated ratio of 1/25? The reason for this discrepancy is that the numerical values of these two probabilities are rounded off to two decimal places.
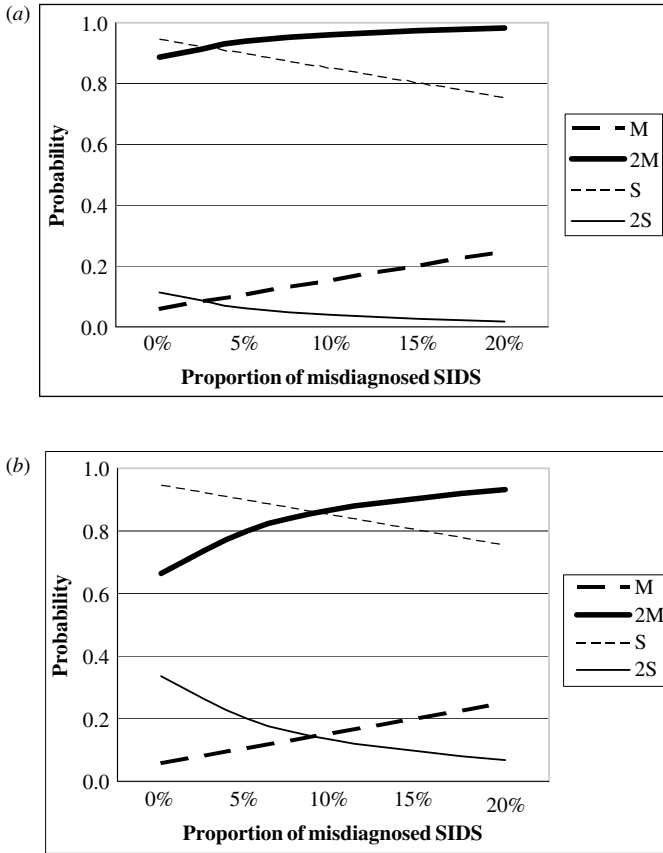
Figure 1. (*a*) Prior probabilities of SIDS and murder (SIDS independence). (*b*) Prior probabilities of SIDS and murder (SIDS dependence).

$p(S_2/S)$ and $p(M_2/M)$. To get a broader perspective on how these probabilities change when values of some of these parameters are altered, take a look at Figures 1(*a*) and 1(*b*). Both figures represent the situations in which SIDS and homicide are the only alternatives, so that $p(S) = 1 - p(M)$, and $p(2S) = 1 - p(2M)$.

Figure 1(*a*) shows how prior probabilities of S, M, 2S and 2M depend on the proportion of prima facie SIDS cases that are in reality covert homicides. The assumption in Figure 1(*a*) is that separate occurrences of SIDS are statistically independent, that is, $p(S_2/S) = p(S_2) = p(S)$.

In Figure 1(*b*), though, the assumption that IH is true is abandoned. In accordance with the earlier review of the literature that disputes IH, it is supposed here that $p(S_2/S) = 4 \times p(S)$.

It is interesting to note that in both Figures 1(*a*) and 1(*b*) the prior probability of *single* SIDS is consistently higher than the prior probability of single homicide (independently of the percentage of misdiagnosed SIDS). In contrast, however, in both figures the prior probability of *double* SIDS is consistently lower than the probability of double homicide. Furthermore, apart from the implausible zero percentage scenario of misclassified SIDS, it transpires that 2M is always *considerably more probable* than 2S. This accords well with the view, expressed by many physicians, forensic scientists and criminologists, that any isolated death classified as SIDS is probably a true SIDS but that two (or more) alleged SIDS deaths in the same family should raise concerns about possible child abuse and would justify more detailed examination.

The inclination to take seriously the possibility of foul play in such situations is opposed by allegedly 'very strong a priori reasons' (Royal Statistical Society [2001]; Royal Statistical Society [2002b]) for supposing that certain genetic or environmental factors predispose some families to SIDS, and that, consequently, an index of suspicion should not be raised much under these conditions. This is not a purely academic issue. Let me give two real-life examples in which reliance on what the Royal Statistical Society called 'very strong reasons' led to a tragic outcome.

In the first case (described in Firstman and Talan [2001], p. 298), a child was many times rushed to the hospital from an early age, with her mother reporting that the baby had stopped breathing and turned blue. A curious thing was that an older sibling had had the same problems but only until the younger child was born, when the symptoms in some miraculous way 'transferred' to her. Some doctors started to suspect that the mother had the so-called 'Munchhausen syndrome by proxy' (MSBP), a disorder where a parent (usually the mother) tries to draw attention to herself by intentionally making her child sick and often eventually killing it.[10] But despite these publicly expressed suspicions, a closer investigation by child welfare authorities and police was blocked by a

---

[10]   MSBP was first described by Roy Meadow, so it was to be expected that with his fall from grace and with 'the mob currently in full pursuit' (Fitzpatrick [2004]) this syndrome would also soon come under a barrage of attacks from vehement, though not necessarily well-informed, critics. And this indeed happened. For instance, although there is a huge specialist literature that analyzes and describes hundreds and hundreds of empirical cases in which MSBP appears to be manifested (which includes even the video recordings of perpetrators caught *in flagrante delicto*), the philosopher Simon Blackburn resolutely asserts that it is 'quite false' that science has discovered that condition. He is equally adamant that there is no reason to believe that mothers ever harm or kill their children merely to gain attention (Blackburn [2005], p. 95). How does he know that? In fact, he gives no justification that would even remotely connect with the debate about MSBP in medical publications. Instead he provides some negative rhetoric, like saying that this syndrome 'merely served to direct power and prestige' to expert witnesses, and led to 'the conviction of many innocent mothers.' There is no argument here, philosophical or otherwise. His intervention only amounted to adding fuel to the fire of the anti-Meadow public sentiment that was already out of control.

senior doctor who had an alternative theory that the child had a heart defect, and so he decided to perform surgery and install a pacemaker. Some time later the child was brought to the hospital unconscious and died soon afterwards. (The pacemaker was functioning properly.) The mother's report about what happened was refuted by a forensic pathologist and she failed the polygraph. But although the district attorney had no doubts about her guilt (like most others involved in the case) he decided not to prosecute because he thought that with the prospect of the senior, highly esteemed doctor defending the SIDS diagnosis he could not convince the jury.

The second case is more dramatic and it involves one of the most widely cited papers about SIDS in the pediatric literature. In 1972 Alfred Steinschneider published an article (Steinschneider [1972]) in the leading journal in the field, in which he analyzed the case of Waneta Hoyt, a woman from New York State, whose three very young children died of unknown causes. Apparently guided by the 'very strong a priori reasons' for believing that certain factors predispose some families to SIDS, Steinschneider hypothesized that the children must have suffered from the same birth abnormality that caused 'apneic episodes' and ultimately death. His suggestion that this disorder might actually be behind many SIDS deaths was soon accepted by a number of medical experts, and the theory stayed at the very center of SIDS research for twenty years. Furthermore, a massive program that included specially designed equipment was established for monitoring the sleep of children who were considered to be at risk. In fact, Waneta Hoyt herself had two more children (beside the three who died), and Steinschneider described in his paper how he put them on 'apnea monitors' in the hospital in order to minimize the probability of the recurrence of SIDS. But despite these measures the two children died one after the other (a year apart), each of them having been only two months old and with the death in each case happening just a day after the child was discharged from the hospital. The diagnosis for both was an apnea-related SIDS. The paper that described the two cases in detail was widely interpreted as strong evidence that SIDS runs in families.

Enter Linda Nelson, a forensic pathologist from Texas. In 1986 she was contacted by a New York State assistant prosecutor for help with another case involving the death of several children in a family. In the course of the conversation about multiple infant deaths she mentioned that she had read Steinschneider's paper and told her colleague that 'the victims in that paper were all homicide victims, in my opinion' (Pinholster [1994], p. 199). This triggered a long and laborious criminal investigation. In Steinschneider's paper, Hoyt's children, Molly and Noah, were referred to only by initials, but this proved sufficient to identify them through death records. Eventually, the first suspicion turned out to be corroborated, and Waneta Hoyt was accused of killing all of her five children. She confessed and explained that she simply

could not stand the children's crying: 'They just kept crying and crying [. . .].
I just picked up Julie and I put her into my arm, in between my arm and my
neck like this [. . .] and I just kept squeezing and squeezing and squeezing. . .'
(Firstman and Talan [1997], p. 435). Hoyt retracted the confession later, but
this was unconvincing (except, predictably, to Steinschneider himself). She
was convicted in 1995 and later died in jail.

The editor of *Pediatrics* wrote in 1997: 'We never should have published
this article [. . .] Huge amounts of time and money have been wasted over the
last 25 years in a useless attempts to ''do something'' for SIDS victims by
monitoring' (Lucey [1997]). Everyone would agree, no doubt, that something
much more valuable than time and money was wasted in the process. Some of
Waneta Hoyt's children might be alive today if more people had kept an open
mind, instead of uncritically trusting their mere hunches and faulty intuitions,
which they mistook for 'a priori very strong reasons.'

5 Likelihoods of Double SIDS and Double Homicide

What about the third probability ratio in Equation (4)? The ratio
p(E/2S)/p(E/2M) tells us about the import of empirical evidence (E) in deciding
between the two hypotheses, 2S and 2M. The question here is whether
the concrete evidence available in the given situation is more likely on the
assumption that 2M is true or that 2S is true. To address that issue, we have to
leave the rarefied area of aprioristic reasoning and descend into the specifics
of a particular controversial case (Sally Clark), which has become something
of a test case for the role of probabilities in this kind of context.[11]

Obviously, before discussing how the two hypotheses relate to empirical
evidence (E), we have first to determine what E actually was in that case.

Helen Joyce describes E in a way that is nothing short of astonishing. She
says that *all* the relevant evidence in the Sally Clark case just reduces to the
fact that 'both children are dead' (Joyce [2002]). From this she concludes that,
trivially, both p(E/2M) and p(E/2S) are equal to 1. (Indeed, the probability
that the children are dead, assuming that they were murdered, is 1. Likewise,
the probability that the children are dead, assuming that they died of SIDS, is
of course 1.)

By reducing the entire empirical evidence (E) to the mere fact that 'the
children are dead,' Joyce is in effect suggesting that Sally Clark was convicted
of murder *without any specific evidence against her being considered*. Joyce even

---

[11]   Let me stress again that, although I will briefly go into some empirical details pertaining to
        Sally Clark's case here, I will not debate the question of her actual innocence or guilt. As I said
        earlier, my exclusive goal is to analyze the statisticians' reasoning. In that spirit, when it comes
        to estimating the import of empirical evidence I will eventually just follow the opinion of one
        of the statisticians (Dawid).

explicitly states that the reason why Sally Clark went to prison was 'because the prosecution argued, and the jury accepted, that lightning does not strike twice' (ibid.). In other words, she says that the guilty verdict was here based exclusively on the prior improbability of two SIDS in the same family.

This is a blatant distortion of what really happened in the court proceedings.[12] In truth, the days of the trial were completely dominated by extensive and detailed discussions of *pathological evidence* of various kinds, while the rarity of double SIDS was just briefly mentioned by Meadow in an aside. The point was not regarded important even by the defense lawyers, who did not think, at the time, that it deserved comments or rebuttal.

What totally discredits Joyce's claim that the jury's decision was based exclusively on their belief that lightning does not strike twice is the fact that the judge actually gave the jury a clear and stern warning that they should *not* rely only on evidence from population statistics: 'I should I think, members of the jury, just sound a note of caution about the statistics. However compelling you may find those statistics to be, we do not convict people in these courts on statistics. It would be a terrible day if that were so.' (Court of Appeal [2000], para. 128)

I am not saying that Meadow's figure of 1 in 73 million did not influence the jury's decision at all. Rather, what I am saying is that Joyce's claim that the specific empirical evidence in the case is exhausted by the fact that 'the children are dead' is a ridiculous misrepresentation of the actual state of affairs.

But what exactly were these other pieces of empirical evidence that went unmentioned by Joyce? Dawid says that 'there was additional medical evidence, including hemorrhages to the children's brains and eyes' ([2002], p. 78). This is still far from satisfactory. The word 'including' indicates that there were yet some other relevant details that Dawid for whatever reason chose not to specify. Clearly, we need all these facts in the open in order to estimate $p(E/2M)$ and $p(E/2S)$, be it even informally and in a very approximate manner. Here are some of these data as they were summarized in the judgment of the Court of Appeal in 2000:[13] (i) inconsistent statements that Sally Clark gave

---

12 Strangely enough, it appears that Joyce's bad advice about how to deal with probabilities in this context is actually endorsed by the Royal Statistical Society, because when Sally Clark is mentioned on the RSS website <www.rss.org.uk/main.asp?page=1225>, there is a direct link to Joyce's paper *and to no other source*. Joyce, in turn, links her paper to the website of the Campaign to free Sally Clark, and she even invites the reader to become involved and to 'write to your MP or to Sally Clark herself' (Joyce [2002]). So, the Royal Statistical Society chose to make its webpage with its pronouncement about probabilities only two clicks away from an Internet site that was entirely devoted to partisan efforts to defend an accused person in a criminal court case. The RSS did not direct its web visitors to other sources of information for a more balanced perspective.

13 I give the snapshot of the relevant evidence at this point in time because the statisticians that I criticize intervened in the case on the basis of what was known *at that stage*. Needless to say, this is not meant to be a full account of the arguments given by both sides, but rather a

to the paramedics and later at the hospital about where one of the babies was when it died,[14] (ii) a child's previous unusual nosebleed unsatisfactorily explained, (iii) a torn frenulum, (iv) extensive fresh bleeding around the spine, (v) hypoxic damage to the brain which occurred at least 3 hours before death, (vi) fracture of the second rib which was some 4 weeks old for which there was no natural explanation, (vii) dislocation of the first rib, (viii) petechial hemorrhages to the eyelid that were acknowledged even by the defense expert to be 'a worrying feature...'.[15]

All these things, especially when taken together, are not what one expects in a legitimate SIDS death. But in an infant homicide, they are not surprising. For if a mother kills her baby, then there is nothing particularly odd if the medical investigation finds post mortem signs of child abuse, like bone fractures and other injuries. In normal situations, however, as the judge in the first appeal said, 'young, immobile infants do not sustain injury without the carer having a credible history as to how the injury was caused.' (Court of Appeal [2000], para. 239) No credible history was offered by Clark's defense.[16]

It might seem that rib fractures could have easily been the result of resuscitation efforts, but this is in fact extremely implausible. Here is the explanation by John L. Emery, a scholar who is held in such a high esteem that he is referred to as 'the master of pediatric pathology' and 'a pediatric pathologist par excellence' (Barness and Debich-Spicer [2005]):

summary of some of the points presented by the prosecution. My aim here, as stated before, is not to argue the guilt or innocence of Sally Clark, but merely to demonstrate that a substantial amount of empirical evidence was considered by the court, and that if taken at face value this evidence would reduce the ratio p(E/2S)/p(E/2M) from Joyce's value of 1. Of course, the defense and prosecution strongly disagreed about the exact probative force that should be attached to that evidence—but a comprehensive consideration of all the arguments made by both sides would lead us too far afield. For a fuller account of the empirical evidence considered at the trial, the reader should consult (Court of Appeal [2000], [2003]).

[14] The incoherence is briefly summarized in a fine article written by two law professors from the London School of Economics: 'She claimed during questioning that one of the babies had died in a baby chair in a position that was not physically possible. Later she refused to answer questions. Then at trial she gave evidence that the child had not died in the chair.' (Nobles and Schiff [2004], p. 236).

[15] Let me repeat once more that I am *not* mentioning these things here in order to argue for Sally Clark's guilt. My point is merely that, when the statisticians want to estimate *overall* probabilities of guilt and innocence by applying the Bayesian inference to this particular case, the only correct way to do this is to also include in their calculations the import of the specific empirical evidence admitted to the court. (Besides, bringing these facts to attention might help to dispel the widespread but completely wrong belief that Clark was convicted *only* on the basis of the very low probability of two SIDS in the same family.)

[16] That these facts indeed have *some* evidential relevance to the issue at hand is illustrated in the following apposite comment: 'The Court of Appeal's conclusion that Sally Clark had had an unfair trial did not (legally) mean that she was factually innocent. And even if the undisclosed evidence did indicate a probable cause of death other than by the infliction of physical injuries, the evidence of such injuries, and the lack of any explanation as to how they could have been sustained, point to unanswered questions.' (Nobles and Schiff [2004], pp. 242–3)

> We and others have gone through the movements of resuscitation on cadavers and have found that it is extremely difficult to fracture ribs in an infant by pressing on the chest or by any of the usual methods of artificial respiration. Fractures of the ribs, however, can be relatively easily produced by abnormal grasping of the child's thorax. The presence of fractures in any site in a child younger than 1 year should be considered as caused by abuse unless proven otherwise. (Emery [1993], p. 1099)

Coming back to the Sally Clark case, it is telling that (according to the judgment of the first Court of Appeal in 2000) even one of the defense's own experts thought that 'there were features in both deaths that gave rise to *very great concern*' (Court of Appeal [2000], para. 77, italics added). Moreover, the defense itself actually 'accepted that there were *worrying* and *unusual* features, but submitted that the evidence amounted to no more than suspicion' (*ibid.*, para. 10, italics added). What do the italicized phrases mean? Well, the most natural interpretation of 'worrying' and 'unusual' is that the features in question were worrying and unusual *for the double SIDS hypothesis*. In other words, these features were easier to explain on the assumption that 2M is true than on the assumption that 2S is true. Or, alternatively, if these features are labeled 'E', then $p(E/2M) \gg p(E/2S)$. Therefore, it appears that even Sally Clark's defense team realized that, given the evidence presented to the court, it would not sit well with the jury to deny the obvious (the existence of suspicious circumstances) and to insist, à la Joyce, that $p(E/2M) = p(E/2S)$.

After it is conceded that $p(E/2M)$ is higher than $p(E/2S)$, the next question is: how much higher is it actually? We need a number for the ratio $p(E/2S)/p(E/2M)$, which could then be plugged into Equation (6) to yield the final answer to our central question (about the ratio of *posterior* probabilities of 2S and 2M). Dawid, confronted with the same need to come up with a numerical estimate, suggests that a reasonable figure for the likelihood ratio is 1/5 (Dawid [2002], p. 78). That is, he proposes that the 'worrying' evidence of the kind that was found in the Sally Clark case is five times more probable if 2M is true than if 2S is true.

In my opinion, there are good reasons to think that Dawid underestimates the difference between the two probabilities. Think about it this way: if it turns out, as it may well do, that the suspicious circumstances of the aforementioned kind are discovered in, say, about 25% of infant homicides (this is a purely hypothetical figure), would you under these circumstances then really expect that, in accordance with Dawid's ratio of 1/5, the disturbing features like babies' broken bones, other strange injuries and the parents' incoherent reports about the death scene would be present in as many as 5% cases of true SIDS? I doubt that you would. If you agree, it is arguable that $p(E/2M)$ is more than five times higher than $p(E/2S)$.

Nevertheless, I do not intend to press this objection. Rather, I will work here with Dawid's ratio, in the spirit of my attempt to evaluate the statisticians' conclusions on the basis of their own premises. Now, with the estimation of the last unknown (the ratio of likelihoods of 2S and 2M), the ground is cleared for the final analysis.

6 Posterior Probabilities of Double SIDS and Double Homicide

If Dawid's proposal of 1/5 for the ratio $p(E/2S)/p(E/2M)$ is substituted into Equation (6), the result is:

$$\frac{p(2S/E)}{p(2M/E)} = 5.6 \times 0.0069 \times 0.2 = 0.008 \qquad (7)$$

This is the probability ratio of the two rival hypotheses (double SIDS and double murder) after *all* the relevant evidence is taken into account. The probability of 2M is 125 times higher than the probability of 2S. Assuming that 2M and 2S are the only two possibilities, the ratio of these two probabilities straightforwardly yields the information about the probabilities of the two hypotheses. All things considered, after the statisticians' reasoning is corrected for numerous errors it turns out that, on their own premises, the statistical probability that Sally Clark killed her two babies would be higher than 0.99 (125/126), whereas the complementary probability that the children died of SIDS would be lower than 0.01 (1/126).[17]

We have to remember, though, that the result in Equation (7) is obtained under two specific assumptions: (a) that the proportion of homicides that are falsely diagnosed as SIDS is 10%, and (b) that IH is true, that is, $p(S_2/S) = p(S_2) = p(S)$. Let us now drop these assumptions and see how the posterior probabilities of 2S and 2M are affected by changes in these two parameters. The more general picture emerges in Figures 2(*a*) and 2(*b*).

Figure 2(*a*) shows the posterior probabilities of 2M and 2S under five different percentage scenarios of misdiagnosed SIDS, but in a world in which IH is true. Figure 2(*b*), on the other hand, represents a world in which IH is false, and in which the probability of SIDS in a family with a previous SIDS case is significantly higher than the probability of SIDS in the general population. In accordance with our earlier discussion, we assume that $p(S_2/S) = 4 \times p(S)$.

Conspicuously, the probability of 2M is *always* much higher than 2S. Moreover, under any of the ten scenarios pictured in the two figures, the

---

[17] Again, although our calculated ratio of the two probabilities, $p(2S/E)/p(2M/E)$, is 1/125 (0.008), it does not come out exactly right if we divide $p(2S/E)$ with $p(2M/E)$: 0.01/0.99 = 1/99. The reason for this is the same as explained in footnote 9: the numerical values for these two probabilities are not exact but are rounded off to two decimal places.

(*a*)

(*b*)
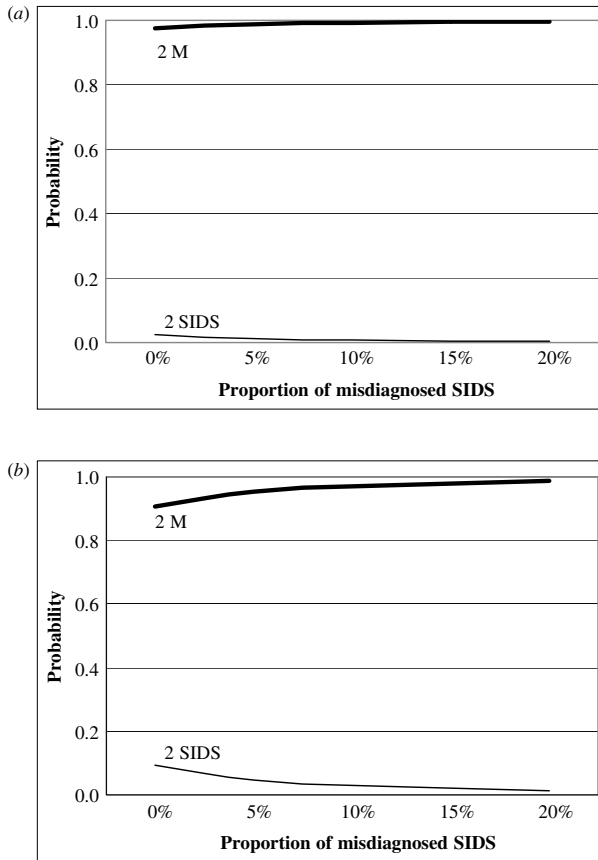
Figure 2. (*a*) Posterior probabilities of double SIDS and double murder (SIDS independence). (*b*) Posterior probabilities of double SIDS and double murder (SIDS dependence).

probability of double murder is greater than 0.9. This is in jarring discrepancy with the estimates of the statisticians. Joyce's final conclusion is that p(2M/E) is 0.33, while Dawid judged it to be as low as 0.04. What is the source of their underestimates?

It seems to me that Joyce's two big mistakes were, first, putting a too high value for p(S₂/S), and second, totally disregarding the case-specific empirical evidence. Nevertheless, despite this fallacious analysis Joyce's explanation of probabilities appears to have been endorsed by the Royal Statistical Society, as well as praised by cosmologist Hermann Bondi in his letter in *Nature* (Bondi [2004]) and by mathematician Barry Lewis in his presidential address as the President of the Mathematical Association (Lewis [2003], pp. 424–5).

Dawid's first major error, in my view, was that he started with a seriously wrong numerical value for p(M). Second, he thought, incorrectly, that if it is permitted to square p(S) to obtain p(2S), then it must also be OK to square p(M) to obtain p(2M). But these two cases are quite different. In estimating the prior probability that Sally Clark's children died of SIDS, squaring p(S) may be controversial but it is certainly not manifestly wrong. (The procedure is actually regarded as perfectly legitimate by many scholars.) In contrast, to do the same when estimating the prior probability that the children's deaths were homicides would fly in the face of basic common sense. Speaking about p(M_2/M), it is grossly implausible to claim that the probability that a child is killed, *given that it is in care of the mother who already killed one child*, is identical to the probability of an infant homicide in the general population.

7 Conclusion

The statisticians have with one voice condemned a particular treatment of probabilities in the Sally Clark case. After proposing their alternative account they have argued that statistical evidence should be presented in the court 'only by appropriately qualified statistical experts' (Royal Statistical Society [2001]) and that 'interpretation of statistical evidence cannot be safely left in the hands of the uninitiated' (Dawid [2005], p. 8). The problem, however, is that the arguments defended by the 'initiated' also suffer from serious flaws and prove on closer inspection to be logically unsound. Therefore, although it is undoubtedly true that court experts, judges and juries could benefit from a better understanding of probabilistic reasoning, this paper shows that even among the 'appropriately qualified statistical experts' there is room for improvement in this respect. *Statistice, cura te ipsum*!

Department of Philosophy
Lingnan University
Hong Kong
sesardic@ln.edu.hk

References

Aitken, C. G. G. and Taroni, F. [2004]: *Statistics and the Evaluation of Evidence for Forensic Scientists*, Chichester: John Wiley.

American Academy of Pediatrics [2000]: 'Changing Concepts of Sudden Infant Death Syndrome: Implications for Infant Sleeping Environment and Sleep Position', *Pediatrics*, **105**, pp. 650–6.

American Academy of Pediatrics [2001]: 'Distinguishing Sudden Infant Death Syndrome from Child Abuse Fatalities', *Pediatrics*, **107**, pp. 437–41.

Bacon, C. [2005]: 'Repeat Sudden Unexpected Infant Deaths', *Lancet*, **365**, p. 113.

Barness, E. G. and Debich-Spicer, D. E. (*eds*). [2005]: *Handbook of Pediatric Autopsy Pathology*, Totowa, NJ: Humana Press.

Batt, J. [2004]: *Stolen Innocence: A Mother's Fight for Justice, The Story of Sally Clark*, London: Ebury Press.

Beckwith, J. B. [1990]: 'Sibling Recurrence Risk of Sudden Infant Death', *Journal of Pediatrics*, **117**, pp. 513–4.

Blackburn, S. [2005]: *Truth: A Guide for the Perplexed*, London: Allen Lane.

Blackmon, L. R., Batton, D. G., Bell, E. F., Engle, W. A., Kanto, W. P., Martin, G. I., Rosenfeld, W. N., Stark, A. R. and Lemons, J. A. [2003]: 'Apnea, Sudden Infant Death Syndrome, and Home Monitoring', *Pediatrics*, **111**, pp. 914–7.

Bondi, H. [2004]: 'Statistics Don't Support Cot-Death Murder Theory', *Nature*, **428**, p. 799.

Brookman, F. [2005]: *Understanding Homicide*, London: Sage.

Brookman, F. and Nolan, J. [2006]: 'The Dark Figure of Infanticide in England and Wales: Complexities of Diagnosis', *Journal of Interpersonal Violence*, **21**, pp. 869–89.

Carpenter, R. G., Waite, A., Coombs, R. C., Daman-Willems, C., McKenzie, A., Huber, J. and Emery, J. L. [2005]: 'Repeat Sudden Unexpected and Unexplained Infant Deaths: Natural or Unnatural?', *Lancet*, **365**, pp. 29–35.

Court of Appeal [2000]: Clark, R v [2000] EWCA Crim 54 (2nd October, 2000), <www.bailii.org/ew/cases/EWCA/Crim/2000/54.html>

Court of Appeal [2003]: Clark, R v [2003] EWCA Crim 1020 (11 April 2003), <www.bailii.org/ew/cases/EWCA/Crim/2003/1020.html>

Dalrymple, T. [2003]: 'Much Maligned', *Daily Telegraph* (December 14th).

Dawid, A. P. [2002]: 'Bayes's Theorem and Weighing Evidence by Juries', in R. Swinburne (*ed.*), *Bayes's Theorem*, Oxford: Oxford University Press.

Dawid, A. P. [2004]: *Probability and Statistics in the Law* (Research report No. 243), Department of Statistical Science, University College London.

Dawid, A. P. [2005]: 'Statistics on Trial', *Significance*, **2**, pp. 6–8.

DiMaio, V. J. and DiMaio, D. [2001]: *Forensic Pathology*, 2nd edition, Boca Raton: CRC Press.

Donnelly, P. [2005]: 'How juries are fooled by statistics', video available online at <www.ted.com/index.php/talks/view/id/67>

Emery, J. L. [1989]: 'Is Sudden Infant Death Syndrome a Diagnosis?', *British Medical Journal*, **299**, p. 1240.

Emery, J. L. [1993]: 'Child Abuse, Sudden Infant Death Syndrome and Unexpected Infant Death', *American Journal of Diseases in Childhood*, **147**, pp. 1097–100.

Firstman, R. and Talan, J. [1997]: *The Death of Innocents: A True Story of Murder, Medicine, and High-Stake Science*, New York: Bantam.

Firstman, R. and Talan, J. [2001]: 'SIDS and Infanticide', in R. W. Byard and H. F. Krous (*eds*), 2001, *Sudden Infant Death Syndrome: Problems, Progress and Possibilities*, London: Arnold.

Fitzpatrick, M. [2004]: '*The Cot Death Controversy*', <www.spiked-online.com/Printable/0000000CA3D8.htm>

Fleming, P. J., Blair, P. S., Bacon, C. and Berry, J. (*eds*). [2000]: *Sudden Unexpected Deaths in Infancy: The CESDI SUDI Studies 1993–1996*, London: The Stationery Office.

Gornall, J. [2005]: 'The Devil You Don't Know', *The Times* (June 24th).

Green, M. [1999]: 'Time to Put "Cot Death" to Bed', *British Medical Journal*, **319**, pp. 697–8.

Guntheroth, W. G., Lohmann, R. and Spiers, P. S. [1990]: 'Risk of Sudden Infant Death Syndrome in Subsequent Siblings', *Journal of Pediatrics*, **116**, pp. 520–4.

Hill, R. [2004]: 'Multiple Sudden Infant Deaths–Coincidence or beyond Coincidence?' *Paediatric and Perinatal Epidemiology*, **18**, pp. 320–6.

Hill, R. [2005]: 'Reflections on the Cot Death Cases', *Significance*, **2**, pp. 13–6.

Home Office [2001]: *Criminal Statistics: England and Wales 2000*, Home Office.

Horton, R. [2005]: 'In Defense of Roy Meadow', *Lancet*, **366** (July 2), pp. 3–5.

Irgens, L. M., Skjaerven, R. and Peterson, D. R. [1984]: 'Prospective Assessment of Recurrence Risk in Sudden Infant Death Syndrome Siblings', *Journal of Pediatrics*, **104**, pp. 349–51.

Irgens, L. M. and Peterson, D. R. [1988]: 'Sudden Infant Death Syndrome and Recurrence in Subsequent Siblings', *Journal of Pediatrics*, **112**, p. 501.

Joyce, H. [2002]: 'Beyond Reasonable Doubt', *Plus*, Issue 21, <plus.maths.org/issue21/features/clark/>.

Keens, T. G. [2002]: *Sudden Infant Death Syndrome*, <www.californiasids.com/UploadedFiles/Forms/SIDS%20Overview.pdf>

King, M. [2004]: *What Fates Impose: Facing Up To Uncertainty (The Eighth British Academy Annual Lecture)*, <www.britac.ac.uk/pubs/src/_pdf/king.pdf>

Levene, S. and Bacon, C. J. [2004]: 'Sudden Unexpected Death and Covert Homicide in Infancy', *Archives of Disease in Childhood*, **89**, pp. 443–7.

Lewis, B. [2003]: 'Taking Perspective', *Mathematical Gazette*, **87**, pp. 418–31.

Lucey, J. F. [1997]: 'Why All Pediatricians Should Read This Book', *Pediatrics*, **100**, p. A77.

Marks, M. N. and Kumar, R. [1993]: 'Infanticide in England and Wales', *Medicine, Science and the Law*, **33**, pp. 329–39.

Marshall, E. [2005]: 'Flawed Statistics in Murder Trial May Cost Expert His Medical License', *Science*, **322**, p. 543.

Matthews, R. [2005]: 'Matt's Stats: Does Faith Have a Prayer in Making a Difference?' *Daily Telegraph* (July 27th).

Meadow, R. [1997]: *ABC of Child Abuse*, 3rd edition, London: BMJ.

Nobles, R. and Schiff, D. [2004]: 'A Story of Miscarriage: Law in the Media', *Journal of Law and Society*, **31**, pp. 221–44.

Office of National Statistics [1998]: *Mortality Statistics: Cause (Review of the Registrar General on Deaths by Cause, Sex and Age, in England and Wales, 1997)*, London: Stationery Office.

Office of National Statistics [2005]: *Homicide Numbers and Rates, 1901–1998*, <www.statistics.gov.uk/StatBase/xsdataset.asp?vlnk=1880&More=Y>

Peterson, D. R., Sabotta, E. E. and Daling, J. R. [1986]: 'Infant Mortality among Subsequent Siblings of Infants Who Died of Sudden Infant Death Syndrome', *Journal of Pediatrics*, **108**, pp. 911–4.

Pinholster, G. [1994]: 'SIDS Paper Triggers a Murder Charge', *Science*, **264**, pp. 198–9.

Royal Statistical Society [2001]: *News Release: Royal Statistical Society Concerned by Issues Raised in Sally Clark Case*, October 23.

Royal Statistical Society [2002a]: 'Report of the Council for the Session 2001–2002', *The Statistician*, **51**, pp. 485–563.

Royal Statistical Society [2002b]: *Letter from the President to the Lord Chancellor Regarding the Use of Statistical Evidence in Court Cases*, January 23.

Steinschneider, A. [1972]: 'Prolonged Apnea and the Sudden Infant Death Syndrome: Clinical and Laboratory Observations', *Pediatrics*, **50**, pp. 646–54.

Sweeney, J. [2000]: *73 Million to One (BBC Radio Five Documentary)*, July 15.

Wilczynski, A. [1997]: *Child Homicide*, London: Greenwich Medical Media.

Wolkind, S., Taylor, E. M., Waite, A. J., Dalton, M. and Emery, J. L. [1993]: 'Recurrence of Unexpected Infant Death', *Acta Paediatrica*, **82**, pp. 873–6.