# Sally Clark – What Went Wrong?

Roger J. Brown, PhD[1]

In 1996 and again in 1998 Sally Clark lost an infant child under 3 months of age to Sudden Infant Death Syndrome (SIDS). Local prosecutors filed murder charges, a jury convicted her, an appeal was lost and she served several years of a life sentence. A second appeal involved testimony by a number of distinguished mathematicians, all of whom spoke of a gross error in probability calculations made by a prosecution witness. The Sally Clark case is fully described at her website http://www.sallyclark.org.uk/.

## I. The variable of interest

It is an artifact of mathematics that careful definitions must be formulated at the outset. In statistics where repeated trials or events are to be aggregated for study in order to draw inference it is important that the elements of the aggregation be as similar as possible. Ideally the similarity should tend toward identical. Less than identical biases the process.

The definition of SIDS is vague. Indeed, the pathology warrants a name which implies that one does not know cause. Families are unique so similarity is very difficult to achieve.

The logic is shown by the alternative. Suppose a group of x infants who died in their sleep from an unknown cause are mixed with a group of y infants who died in their sleep from carbon monoxide poisoning. This is an obvious error and no one would study the group composed of the combination x and y. Now suppose that "unknown cause x" can include two or two hundred different causes, all different but all unknown. The same error occurs. Some coroners prefer "unascertained" rather than SIDS because the term SIDS is not sufficiently definite.

The root of the problem is found in the underlying basis for all of probability, which is a simple division problem:

$$\frac{\text{everything that DOES happen}}{\text{everything that CAN happen}} \leq 1$$

In order for a rational statistical discussion to take place it is necessary that (a) those items constituting "everything" be carefully defined; and (b) the individual elements that collectively comprise "everything" are as similar as possible.

To continue along the path of absurdity, if one wanted to study car accidents one would include driving and roadways in the universe of possibilities. It would be questionable at best to include aircraft accidents (even if there is the rare case of a collision between a car and a plane).

[1]  The author wishes to thank Ray Hill and Andreas Lauschke for very helpful comments.

Hence, to provide sound mathematical guidance, at the granular level we want a clear and clean definition of substantially similar events. SIDS, depending on who you ask, at least partially or grossly fails this test.

## II. The role of jurisprudence

The law is interested in civil order and justice for those in violation of societal norms. Suspected infant homicides lend themselves to a different kind of investigation. Modern science can determine if some infant deaths are not from natural causes.

Taking infant deaths from all causes one encounters a subset of those deaths for which the cause is unknown. Further dividing those into infanticide and SIDS, the proportions may be compared to arrive at relative probabilities. A British study provides this insight:[2]

The chance of choosing an infant that died from SIDS is about 1 in 1303

$$\frac{SIDS\ deaths}{live\ births} = \frac{363}{472{,}823} = .000767 \Rightarrow 1\ of\ every\ 1303$$

The chance of choosing an infant that was murdered is about 1 in 21700

$$\frac{infant\ homicides}{live\ births} = \frac{30}{650{,}000} = .0000461538 \Rightarrow 1\ of\ every\ 21{,}667$$

Relative likelihood of infanticide and SIDS shows SIDS nearly 17 times more likely than homicide

$$\frac{650{,}000/30}{472{,}823/363} = 16.634$$

There are several problems with these calculations. One is that while we are not told what constitutes an "infant", the time period during which the number of births and deaths occurred likely has children from prior periods "aging out" of the sample or children whose infancy continues beyond the sample period. The problem is that birth and death occur at specific points in time and "infancy" occurs over a span of time so some of the deaths recorded likely affected children not in the sample of live births. The second problem is that these numbers are valid only for the period studied. Advances in medicine, parental education and other factors improve survivability over time at all ages in most industrialized societies. Finally, each year medicine identifies more specific causes so that a death which might have been called SIDS in a prior year is death from a known factor this year.

[2] Fleming, P., Bacon, C., Blair, P. and Berry, P.J. (2000) *Sudden Unexpected Deaths in Infancy, the CESDI Studies 1993-1996.* London: Stationary office

## III. The notion of independence

The important idea of independence is best described by consecutive flips of a fair coin. Since the coin has no memory, when flipped a second time it is entirely and completely uninfluenced by the outcome of the prior flip. A fair coin has a .5 (50%) chance of coming up heads. Two flips of that coin have a .5 x .5 = .25 chance of BOTH coming up heads. Three flips have a .5 x.5 x .5 = .125 chance of all three coming up heads, etc.

Because probability is always expressed as a decimal number between 0 and 1, repeatedly multiplying very small numbers always results in even smaller numbers, in the limit approaching zero as in Table 1 below.

| ⌗ flips | Probability of all heads | Decimal Form |
|---|---|---|
| 1 | $\frac{1}{2}$ | 0.5 |
| 2 | $\frac{1}{4}$ | 0.25 |
| 3 | $\frac{1}{8}$ | 0.125 |
| 4 | $\frac{1}{16}$ | 0.0625 |
| 5 | $\frac{1}{32}$ | 0.03125 |
| 6 | $\frac{1}{64}$ | 0.015625 |
| 7 | $\frac{1}{128}$ | 0.0078125 |
| 8 | $\frac{1}{256}$ | 0.00390625 |
| 9 | $\frac{1}{512}$ | 0.00195313 |
| 10 | $\frac{1}{1024}$ | 0.000976563 |

Table 1

The multiplication of probabilities as a way to decide the probability of multiple independent events is very unforgiving. Independence must be as stark and unquestionable as the separate flips of the same coin. It is not easy for this assumption to hold. There are any number examples, ranging from the trivial to the consequential, that might illustrate this. Here is one. Suppose you and I live a mile apart and each have an elm tree growing in our yards. One might assume that those trees are independent. But perhaps they were bought from the same nursery. Or suppose yours was grown from an acorn which fell off mine and carried by a bird that dropped it on your lawn. Nature is very interconnected. The reason inanimate objects like coins and dice are often used to illustrate independence is because the physical realm offers a few very good contexts for independence. Once you introduce biology and human behavior the armor of independence quickly melts away.

The prosecutors, through their expert Roy Meadow, introduced a calculation that, in order to be correct, *required* independence. According to Meadow the chances of a single death from SIDS was one in 8543 (even this was either wrong or prejudiced and was later shown to be closer to 1 in 1303, see Section II above). He improperly reasoned that the probability of two deaths was the product of 1/8543 times itself, a number approaching one in 73 million.

$$\frac{1}{8543} * \frac{1}{8543} = \frac{1}{72{,}982{,}849}$$

This demands independence in the events. How can the deaths of two children of the same parents living in the same house (albeit not at the same time) EVER be independent?

IV. <u>The idea of conditional probability</u>

To better understand the problem and a more accurate result requires an appreciation of the idea of conditional probability. This, in computer parlance, may be thought of as a sort of an "If - Then" statement. A more colorful portrayal is the story about the man who heard that there was only a one in a million chance that a bomb would be on a plane. Since he knew about independence he figured there was a one in a trillion (one million times one million) chance of there being two bombs on board. He therefore always packed a bomb with him whenever he flew.[3]

**Conditional probability and data**

One must be very careful when applying data to individual cases. An observation about mortality tables applies here. Insurance companies can take 100,000 35-year old men who are alive on January 1st and tell you EXACTLY how many will be dead by Dec 31. But they cannot tell you if you are one of the unlucky ones.

Does this make data useless? Not at all. It just argues for its cautious use.

Conditional probability is a very useful concept. An example from the natural sciences illustrates. Suppose I have a sample of the height of each of 5400 males between the age of 2 and 19, inclusive. From that you wish me to estimate the age of your 17 year old son. With only the data I have described I can add up all the individual heights and divide by 5400 and get an average height of the entire group of 56.49 inches. This is known as "the unconditioned mean" (average and mean are equivalent terms) or the expectation (another term that - with some technical differences - is the same as average) of the height of an individual drawn randomly from the sample.

But, you object, your son is nearing six feet tall! What good is my data if I cannot be any more accurate than that? Ah, I reply, the difference is that my data does not know how old anyone is, only that there is no one over age 19 or under age 2 and you KNOW the height of your son.[4]

Now let's change the situation in a meaningful way. Let's ask all 5400 subjects to report their ages. Let us assume for simplicity that the entire sample of 5400 is composed of 18 age groups, each containing 300 subjects. I can now estimate the conditional mean of the sample which is conditioned upon the subjects' age. Using only the 300 17 year-olds I add up their heights, divide

[3] For those who may not find that hilarious, the true odds would be those arising from the chance that someone else brought a bomb on board *given that* a bomb was already on board, an entirely different calculation.
[4] There is a second problem. The average is a linear operator, essentially assuming that each year produces the same amount of growth something that is manifestly not true in human beings.

by 300 and arrive at 69.27 inches. The difference between the unconditional mean and the conditional mean, about 13 inches, is meaningful, and closer to your son's height.

Returning to the Sally Clark case this insight reveals a number of issues. First, the way the expert, Meadow, arrived at 8543 as the denominator in his odds calculation (recall this is only for a single SIDS death) was to screen for affluent, non-smoking families in which the incidence of SIDS death is known to be less. Using the entire sample from which the subset was drawn the chance of any random child dying of SIDS is 1 in 1303 (see Section II above for calculation). His screening, in essence, "meddled in the randomness", cherry-picking a favorable subset (just as I did by selecting only 17 year olds above) that improved his argument by claiming that odds were conditioned upon the absence of two characteristics (smoking and deprivation) known to increase the number of SIDS deaths. This increased the size of the denominator and ballooned the result of multiplying it by itself (the act of which compounded the mistake by ignoring dependence). Missed by prosecution, the defense, the trial judge and the first appeal court, not to mention the jury, was the fact that the same factors (smoking and deprivation) which lower the odds for SIDS also lower the odds for homicide. In balancing the way in which evidence is presented it is appropriate to adjust both sides, not just one, of any comparison to maintain parity.

The second issue has to do with a centuries-old famous result in probability theory by Thomas Bayes.

**Bayesian Conditional probability**

The reason my conditional expectation of your son's height (69.27 inches) is more accurate than my unconditional expectation (56.49 inches) is *that I have added information* to the expectation when I come into possession of the subject's age. Adding information, in a legal setting may be viewed as "new evidence" and is accorded the same importance. By adding new evidence we change the probabilities.

Rather than selecting only 17 year olds, there is a formula for measuring probability using all the data which reaches the same conclusion. For this we need some notation and a definition. The probability of an event occurring *given that* some other event has occurred is expressed as P(A|B), in words: "Probability of A given B." With this we state Bayes Theorem:

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

In words this says that the Probability of A given B is equal to the product of the Probability of B given A times the Probability of A, which product is then divided by the Probability of B.[5] While this may seem mind-numbingly complex it is really very simple, involving nothing more repeated application of our simple division problem in Section I.

---

[5] It is worth pointing out that the product in the numerator is valid because the events B and A are independent, notwithstanding the fact that A occurred prior to B.

Returning to the time before I knew anyone's age, the best I could do in predicting that someone I might draw from the whole crowd was more than 70 inches tall was to find all those subjects (there were 146) and compare that number to the entire group

$$P(A) = P(Height \geq 70\ inches) = \frac{number\ of\ males\ who\ ARE\ \geq 70\ inches}{number\ of\ males\ who\ CAN\ be\ \geq 70\ inches} = \frac{146}{5400}$$

The fraction you see at the end may, of course, be converted to decimal (.0207) or percent (2.07%) but for ease of exposition let's leave it in the equation as a fraction for now. Our unconditioned expectation, based only on the data of heights for 5400 males is that just over 2% are 70 inches or more.

Now we add the age factor. After revisiting our group and collecting their ages we learn that:

$$P(B) = P(Age = 17\ years) = \frac{number\ of\ males\ who\ ARE\ 17\ years\ old}{number\ of\ males\ who\ CAN\ be\ 17\ years\ old} = \frac{300}{5400}$$

This is the unconditional expectation of picking a 17 year old (about 5.6%). Now we need a conditional probability, that of finding a 17 year old that is at least 70 inches tall:

$$P(B|A) = P(Age = 17\ years|Height \geq 70\ inches) = \frac{number\ of\ 17\ year\ olds\ who\ ARE\ \geq 70\ inches}{number\ of\ 17\ year\ olds\ who\ CAN\ be\ \geq 70\ inches} = \frac{2}{300}$$

The rest is very simple. For the Probability that a male is over 70 inches tall, given that he is 17 years of age, which is P(A|B) from above, we have merely three fractions:

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)} = \frac{\frac{2}{300} * \frac{146}{5400}}{\frac{300}{5400}} = \frac{73}{22{,}500} = .00324 = .324\%$$

This outcome is intuitively reasonable. Our data (I looked) contains no males under the age of 17 or actually age 18 who are more than 70 inches tall. There are only 146 total who are more than 70 inches tall, two are 17 and all the rest are 19 years of age. So there is a very small chance (about a third of one percent) of selecting a random male who is both 17 years of age and 70 inches tall. It is also easy to rule out six foot tall four year olds. When the "2" in the numerator above is zero (meaning we have none) the probability is zero as equation produces is zero

We can further visualize this graphically. In Figure 1 suppose that the yellow disk is the universe of our entire group of 5400 males. We know that there are 300 in each age category. We will start with 300 males who are 18 years of age, represented by the blue circle. Notice they do not overlap because, as we found from inspecting the data, there are no 18 year olds who have reached 70 inches tall.

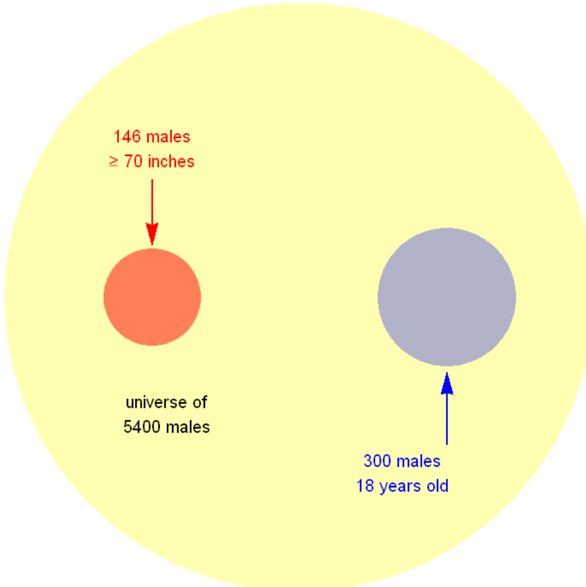


Figure 1

However, for the subset of 17 year olds we have overlap because two of them are 70 inches or more. The graphic changes in Figure 2. The intersection of the blue and the red produces our probability (.00324=.324%) that a random draw from the entire group will produce a male who is both 17 years of age and 70 inches tall.

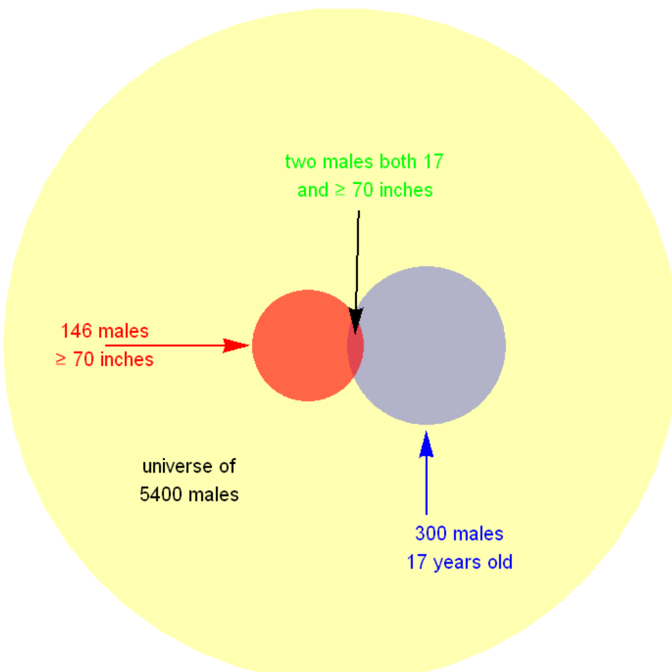


Figure 2

Well documented in the various papers written about the Sally Clark case[6] are the equivalent calculations leading to the conclusion that the probability of her innocence was closer to one in nine than one in 73 million. Meadow's error is one of breathtaking proportions.[7]

## V. The Prosecutor's Fallacy

Full descriptions and illustrations of the Prosecutor's Fallacy may be found in many places. It was used successfully (and tragically) on the Sally Clark jury. The essence of this fallacy can be summed up simply as: Just because a rare event happens does not mean you caused it.

## VI. Summary

In all that has been written about the Sally Clark case, cases similar and the troubling SIDS issue in general, one wonders if statistics belong in the courtroom at all. There usually is one defendant. Choosing that defendant out of a group and deciding her guilt or innocence on the basis of mathematics will strike some as crude. The prosecution may wish the defendant to be so representative of the perpetrator as to make it impossible not to convict. But the defendant is, in fact, a sample of one. The law wants to treat individuals as individuals. The Law of Large numbers does not do this well.

The key, if mathematics and statistics are to be used at all, is to make good calculations based on sound mathematics and interpret it carefully and correctly. This is not an easy matter.

In an adversarial setting it is incumbent on both parties to seek the truth much like George Bernard Shaw had in mind when he formulated his Theory of Legal Procedure.[8] In California's Family Court, when real property valuation is necessary, it is required that the warring parties agree on a single appraiser, the cost of which they share equally. Each side then hires their own expert (the author has served in this role several times) to critique the appraisal. There is also a growing use of a discovery practice, dubbed "Dueling Experts" in which both experts are present during deposition and debate openly on the record the merits of each other's opinion.

Finally, attorneys need to recognize that mathematics follows them into the courtroom. Sound logic and good reasoning are the foundation of mathematics *and* correct decisions. While courtrooms can be scenes of histrionics, when judgment is rendered calmer heads must prevail. Guiding those calmer heads with mathematics is an excellent start.

---

[6] Cf. Hill, R. (2004) Cot death or murder – weighing the probabilities. *Developmental Physiology Conference*, June 2002 and Hill, R. (2004) Multiple sudden infant deaths – coincidence or beyond coincidence? *Paediatric and Perinatal Epidemiology*, 18, 320-326. One of Ms. Clark's attorneys wrote a book about the case. Batt, J. (2004) <u>Stolen Innocence; the Story of Sally Clark</u>. London: Ebury

[7] One might well ask why not show the correct way to make Meadow's calculation? It is easier to describe what went wrong using an example based on height and age because those variables may be measured and the result uncontroversial. The calculation using data from the CESDI study requires assumptions about the degree of dependency arising from genetic and environmental factors, about which debate may remain. Nonetheless, the assumptions need not be heroic to leave the original calculation stunningly in error.

[8] For the few remaining souls unfamiliar with this quip, I repeat it once more: "The Theory of Legal Procedure is that if you set two liars to expose one another, the truth will emerge"