Volume 21, Number 2, April/May 2015

Institute for Health and Society, Medical College of Wisconsin

MEDICAL COLLEGE OF WISCONSIN

Datum

NEWSLETTER
Division of Biostatistics

ADVANCING HEALTH THROUGH RESEARCH AND DISCOVERY CTSI

The Importance of Being Correctly Conditioned (and the Sally Clark Case)

Anjishnu Banerjee, PhD, Division of Biostatistics, MCW

Conditional probability: The concept of conditional probability is a fundamental underlying principle, pervading through a lot of statistical methods, but is often ignored in practical implementations. In this article we present the Sally Clark case, a criminal trial attracting a lot of publicity in its day, where misuse of statistics lead to disastrous consequences. Along the way, we present a few concepts of conditional probability and independence.

The Sally Clark case: Sally Clark was a young British solicitor, who in 1999, was convicted of murdering her two infant sons, but was later acquitted in 2003, on second appeal. Her first son died in his infancy, barely a few weeks old, in September 1996, and the second son, died under similar circumstances, in December, 1998, about 8 weeks old. Arrested and convicted of murder, the conviction primarily relied on statistical evidence presented by a prominent pediatrician, Sir Roy Meadow, who told the court that the chances of two cot deaths (SIDS is the formal name for cot deaths now) was "one-in-seventy three million" for an affluent non-smoking family like Sally Clark's. This estimate was controversial at many levels (even though it was not questioned when presented in court), but there were at least two fundamental flaws with it:

a) The chance of "one-in-seventy three million" for there being two cot deaths even if reasonable, does not equate to the probability of Sally Clark being innocent as "one-in-seventy three million", as was reported in leading British newspapers then. Given the two cases of infant deaths, we should weigh up the relative likelihoods of competing explanations – (i) two successive SIDS, (ii)Sally Clark being guilty of double murder, (iii) other explanations like one murder and one SIDS. Even though double SIDS was unlikely, double infant homicide was unlikelier still, so there would have been a large chance that Sally Clark would have been innocent. Much later, a professor of mathematics and statistics, Ray Hill estimated the odds of Sally Clark being guilty as between 4.5 to 1 and 9 to 1 against. We elucidate this concept in our explanation of condition probability later.

b) The figure of “one-in-seventy three million” was arrived at by squaring “one-in-8353”, where “one-in-8353” is the chance that there is one SIDS in a family like Sally Clark’s. This “one-in-8353” estimate was highly dubious, but even if true, squaring it to arrive at the “one-in-seventy three million” figure meant that we were treating the deaths as independent events. There may well be hidden environmental and genetic factors at play, so that one SIDS case in a family makes another SIDS case highly likely, as later research has shown – in fact the chances are increased by factors between 5 to 10. The squaring of the estimate was quite nonsensical and certainly very flawed science.

The aftermath of the case was not a pleasant one – Sally Clark, when in prison, was frequent target of other female prisoners, possibly due to her history as a solicitor and her conviction as a child-killer. After her acquittal she developed serious mental health issues and died of alcohol abuse in 2007. With her acquittal, the attorney general ordered the review of hundreds of similar cases, when several other women were acquitted. Sir Meadow was removed from the British General Medical Council, guilty of serious professional misconduct, but reinstated later.

<u>Conditional Probability</u>: In simple terms, conditional probability is probability that a certain event occurs, given that some other event has taken place. In the Sally Clark case, we were interested in the probability that Sally Clark was innocent given the event that two of her infant children had died. To illustrate the estimation of conditional probability, we make up a couple of hypothetical numbers: 1) Suppose that a diagnostic test for a certain type of rare cancer was 99% accurate and 2) Suppose that the prevalence of the cancer in the general population was 1%. Now we suppose a random person tests positive with the diagnostic test, what is the probability that she/he actually has cancer? This is not 99%, as is often incorrectly estimated – because we’ve not taken into account the rarity of the disease here.

Suppose we have 10000 people in our population of interest. If the estimated rates of occurrence are correct, we’d have about 100 people with the cancer. If the test was administered on all of these 100 people, and if the test was indeed 99% accurate, we’d find about 99 of these 100 people testing positive. But in the remaining population of 9900 on healthy people, we’d probably find 99 people incorrectly testing positive – these are often dubbed “false positives”. Correctly accounting for the rarity of the disease, we’d find that the chances of person having the cancer, given that she/he’d tested positive to be just 50% and not 99% !! These calculations are also illustrated in the table below:

	Test positive	Test negative	Totals
Have Cancer	99	1	100
Healthy	99	9801	9900
Totals	198	9802	Grand Total: 10000

Therefore, probability(have cancer, given a positive test) = 99/198 = 0.5 or 50%

Similar arguments were used by Professor Hill to estimate the odds in the Sally Clark case.

<u>Conclusion:</u> Ignoring conditioning information can lead to serious statistical misinformation and this is even more exacerbated in case of rare events. Whenever needed or present, estimation of proportions or rates between groups should account for conditional information. There are many ways of including conditional information in our statistical analysis – for example, inclusion of so-called confounders produces conditional estimates and helps obtain more reliable estimates.