"Effects of beer on breast fed infants"

- **How would you - a priori - have decided the sample size for this study?**

  Ask experts what would be an important reduction Δ in amount consumed. Need some idea of the SD of a within-child difference in consumption from a four hour period in one day to a four hour period in another day. Could do a pilot study with say 10 children measured [without any intervention] on two different days and get the SD of the 10 differences. Then use sample size formula for 1 sample t-test with whatever alpha and beta decided upon.

- **Do you have a way to reconstruct the SD of the 11 within-pair differences? If yes, explain how; if not, why not?**

  We know t = dbar / SD(11 differences) /√11 = 2.47. From xbar1 and xbar2 we have that dbar = 193.1 – 149.5 = 43.6 so we can work back to SD(11 differences) = 43.6 • √11 / 2.47 = 59.

- **What do you think the ±18.4 and ±13.1 are? What are other possibilities and why do you tend to rule them out?**

  Just by their size they are too small to be SD's measuring the variation across 11 children in one session (children are not that homogeneous). Working back from the SD of 59: We know that the SD of a difference is roughly √2 times the SD of each individual set if the 2 sets of observations are uncorrelated. Thus if SD(difference) = 59 = √2 SD(individual observations in one session) we would have SD(across individuals at one session) = 42. If there was a correlation r between the 2 sessions ie if a child who was above the average of the 11 on one session tended to be above/below the average of its group on the other session, then the SD(11 differences) = 59 would equal SD(one session) • √2 • √(1–r). But there is no sensible r such that we could get an SD of 59 from SDs of 13 or 18.

  So the 13.1 and 19.4 must be SE's or 2 SE's of the mean at each session. Since one wouldn't expect that r is very large (especially if children were all the same age), one would guess that the SD of 42 or so in a session is not that far off, and if we divide the 42 divided by √11 to get a SEM, it is not be too far from the 13 and 18 reported

- **Are you comfortable with the statistical analysis performed? List 2 other tests that were available to the authors.**

  One could question use of t-test with such small n=11 where we would are unable -- even from the raw data -- to check normality and would have to rely on our expert judgement.
  So instead of relying on the t-test and its uncheckable assumptions, we could use the sign test or the signed rank test (nboth nonparametric)

- **In the last para., why are the authors careful about their inferences?**

  The hypothesis is about milk production but the study measured milk consumption, and did so in only one 4 hour session for each condition. There is the question of blinding, of long term behaviour, of what the mechanism is, etc.

"Thornton Wilder's original design"

- Suppose, before funding him, the funding agency had asked Brother Juniper to document how reproducible his ratings (and overall index) were. Suppose he had consulted you. What data would you have advised him to assemble to answer this question and what data summaries/presentations would you have recommended?

  Would want to see how he would do if asked to re-rate subjects on his scale. Also, would want to see if others understand and can reproduce it. This doesn't even get into the question of the validity of the scale; the first step is usually to see if it is reproducible. I would advise him to use Bland and Altman article ***Lancet 1986 1 307-310*** *(or CV or something quite descriptive) to present results of intra-rater and inter-rater reproducibility.*

- Suppose the pestilence left 9 survivors for every 1 it carried off, and that his budget and other constraints only allowed him to study a total sample size (cases and controls) of only 30 "souls" ("subjects"). Why would Brother Juniper use samples of 15 and 15 rather than say 27 and 3?

  If the variation in the two populations is the same, the SE of the estimated difference is proportional to $\sqrt{(1/n_1 + 1/n_2)}$ *and this is minimised by having* $n_1=n_2$*; a 3:27 is not efficient.*

- Translate the last sentence of the passage from the book ("He added up ... ") into the way one might report the results today (you do not have to use a ratio).

  Total(victims) = 5• Total(Survivors) or xbar(victims) = 5 • xbar(Survivors) .Of course today this would be accompanied by a t-test and a P-value. How about this answer from one student?
  THE DEAD WERE 5 TIMES MORE LIKELY TO SURVIVE THAN THE SURVIVORS

- If he submitted his report to the journal Religio-Epidemiology, the reviewers might had concerns about study design, data analysis, and interpretation. What issues of a more statistical nature might they have raised?

  blinding; sample size; power; reproducibility; statistical significance; clinical significance;

- Not all statistical tests/procedures available today were available in Brother Juniper's time. What relevant statistical tests would have been available to Brother Juniper if he were analyzing his data today?
  t-test for independent samples (1910 or so); rank sum test (1940's)