Reviewer's report

Title: Sample size requirements for case-control study designs

Authors:

Michael D deB Edwardes (michael.edwardes@clinepi.mcgill.ca)

Version: 3 Date: 26 Sep 2001

Reviewer: Dr Jesse Berlin

Level of interest: not specified

Advice on publication: Other (see below)

1. Conceptually, the proposed program seems like a great idea, as it provides in one place a way of comparing sample size estimates obtained from various approaches to design. This is the kind of thing collaborative statisticians are asked to do all the time, so making this program available publicly would be of great potential value.

I should emphasize that I have not verified the correctness of the mathematics. However, this is an area that has been well-explored and the math should be fairly straightforward.

2. The real challenge with an article like this is how to approach the presentation. From my own perspective, treating the paper as a kind of introduction to the program and a tutorial would make it the most useful. Striking a balance between this didactic goal and how much detail to provide on the math is always tricky. It is also a challenge to describe the data requirements for the sample size calculations, although it is easier to do this in the text of the article than in the body of the program.

Having said all that, my own approach to revising the paper would be to focus on the use of the program and to move the details of the algebra to an appendix. Depending on what the editors think of as the focus of the journal, one could argue for the reverse, i.e., giving the math in the paper and including the use of the program as an appendix, which is what you now do. I could be convinced either way, but in any case, I believe what is needed for the program tutorial is a detailed worked example. You pretty much have that now, but I would be even more generous with the inclusion of tables explaining everything. For example, I would include the tables relating exposure to the confounder early on, in the body of the text (or of an appendix), with a fuller explanation of what the hypothetical study is about and what the numbers mean. I'll have more specific comments below on the example you use.

3. Off the top of my head, the only other program I know of that does the adjusted sample size calculations is EGRET SIZ. I'm not even sure it's available anymore. If it is, you might consider comparing your results with results produced by EGRET. In any case, as a statistical paper, a missing element of this paper is some demonstration of the validity of the sample size calculations. Again, as I noted above, the extensions you present to previous results may be trivial enough that this is not an issue. However, validity could be demonstrated through some sort of simulation. That may be beyond the scope of this journal and this paper, but perhaps you could provide more explicit references that address this.
4. As a general comment, it might be worth noting in your discussion that a difficulty with using the program is the detail of the data requirements. One needs to know the expected relationship between exposure and the confounder and between outcome and confounder. This can sometimes come from pilot data or from previous studies, but some comment on this issue seems warranted. You might also note that this approach can be used as kind of sensitivity analysis, looking at how sample size requirements change with various assumptions about the strength of the associations with the potential confounder.

5. One other general question had to do with the interaction. I wasn't clear about whether you are determining a sample size for detecting the interaction itself, or whether you are still interested in estimating the main effect of exposure in the presence of an exposure X confounder interaction. If you are suggesting the latter, would you be fitting a model that includes the interaction term, or just doing a stratified analysis of some kind and still coming up with a single summary estimate of the exposure effect? I would note that many epidemiologists would object, some strongly, to estimating an overall exposure effect in the presence of a known effect modifier. Such an overall effect could be considered by some to be meaningless. I don't necessarily share that view, but you may be subject to some harsh criticism. I'm sure you know all of this. I would suggest that you at least include some kind of disclaimer that admits that not everyone thinks this is a good idea, but we wanted to allow for it.

Specific Comments

1. MINOR POINT page 1, Section 2: Here you say that F is "the ratio of cases to controls (i.e., 1:F)." It is clear later on that F is really the number of controls divided by the number of cases, or put more simply, the number of controls per case. The 1:F does make it more obvious, at least in retrospect once you know the right answer, but some clarification would be helpful here.

2. Section 3: As I noted above, you may want to consider de-emphasizing this section and increasing the emphasis on an example. I don't see where G' is defined initially - you say here that it's definition needs to be modified, but modified from what? For the definitions of the weights, I'm not seeing the connections you are making. For example, W1i is defined as "the proportion of controls". The proportion relative to what? Are you really getting the joint probability of being a control at a given level of the confounder?

Some clarification is needed.

3. MINOR POINT: First paragraph of section 5: Would the unconditional analysis that accounts for the stratification be a stratified analysis (e.g., Mantel-Haenszel or Woolf-type estimate) or a logistic regression with indicator variables?

In the same paragraph, you mention the possibility of joint stratification on more than one confounder. It might be helpful to limit the example to 2 confounders, each with 2 levels, and point out that this would lead to 4 strata. (You could make it 3 confounders, each at 2 levels, leading to 8 strata. The point is to be a little more explicit.)

4. MINOR POINT: On page 4, section 7, for input you define pi as the probability of being in level i of the confounder. On my first reading, it was not clear whether that was for controls only or for controls and cases? Presumably, for a matched study, the proportion at a given level of the confounder would be the same for cases and controls. On rereading, it becomes clear that the DESIGN program is limited to matched studies, so perhaps we just need a small reminder in the introductory paragraph to this section
that these are category-matched studies. Just a suggestion.

5. Do you plan to provide the programs on request? Is there a link to a website where the programs can be downloaded? Would the journal set up such a link on their own site? It seems to me that the electronic format of the journal provides a unique opportunity to make the actual program available. It is another question whether you want to make just an executable version available or also make the code available for those who want to make modifications. Is the STATLIB link what accomplishes all of this? (Forgive my ignorance.)

6. The example you provide seems to be rather extreme in terms of the difference in sample size requirements between the adjusted and matched analyses. Presumably, this remarkable gain in efficiency is related to the large magnitude of the association between the confounder and outcome. The values of $R$ go to 28.5 for the highest level of the confounder. I don't know that I've ever seen a confounder with that large an odds ratio, but if I have, it has not been that common. Are these numbers derived from an actual (real-life) example? Ideally, I would like to see an example based on some real data, where the associations with the confounder can be documented somehow. At a minimum, an example (which could be a second example), with a less extreme confounder-disease association should be included.

7. In the printout from the program, the text says "SAMPLE SIZE REQUIRED TO CONFIRM THAT AN ODDS RATIO (OBSERVED) = 2.00, IS NOT EQUAL TO ONE,...". Presumably, these power calculations (like any others) are based on the true underlying odds ratio, and not on the observed odds ratio. I'm sure you know this, but suggest you modify the program accordingly.

**Competing interests:**

None declared.