Reviewer's report

Title: The National Women's Health Study: methods and description of the study population

Version: 1 Date: 17 March 2004

Reviewer: Pauline Mendola

Reviewer's report:

Is the question posed by the authors new and well defined?

This is an area of great scientific interest and in most areas of the world, reproductive health data are not routinely captured outside of clinical settings. The lack of non-clinical data can lead to potentially biased inferences on common health events such as spontaneous abortion that are not always medically attended. The authors proposed to address this lack of population health information by using a mail survey.

Are the methods appropriate and well described and are sufficient details provided to replicate the work?

A population-based mail survey is an appropriate method to get at this question but the authors have made a series of non-standard decisions along the way that make the study impossible to replicate, but more importantly, make the generalization of the data to the underlying population suspect. Here are some of the problems with the study design and some potential ways the authors might address them:

Electoral roles were used as the first stage sampling frame. Presumably, not all women are included in electoral roles (i.e., non citizens?) but is registration a mandatory function or subject to some compliance or volunteer registration? A description of who is not covered in the sampling frame would be helpful.

Age restriction is a second stage. First, I am not sure that age restriction was necessary and age subgroups could have been easily handled in the analysis phase of the study. Reproductive events are well recalled (as the authors note) and the restriction brings in many problems that could have been avoided. It seems that age is not part of the electoral role data (or that would have been used to select the sample), but choosing to restrict the sample on the basis of the prevalence of certain given names is frankly impossible to interpret without a lot more supporting information. Is there some reference or series of references to describe the validity of this process? Is there a “set” of names that are evaluated or are all names selected under a probability scenario? For example, it is not clear how common names (e.g., Jane, etc.) would be handled. And, most importantly for the internal validity of this study – did this scheme work? Was a random sample of non-selected women contacted or traced to find out the sensitivity and specificity of this process for identifying women under age 55? Approximately 10% of the sample selected was ineligible (>20 % of respondents), and sensitivity and specificity of this restriction process is not a trivial issue.

Moving from the stage one questionnaire to the stage two questionnaire, the authors have an expected level of non-response, poor contact information, and refusals. All expected but the impact on the relation of the sample to the underlying population should be discussed in more detail. A potentially bigger problem is with the exclusion of women who had not attempted pregnancy. More information is needed on how they excluded women who had never attempted pregnancy. Many women have long periods of unprotected intercourse without achieving pregnancy but they were not
“trying” to get pregnant. If this question was not clearly addressed, it is likely that the data will underestimate primary infertility in the population. Given the low prevalence of primary infertility found in the study – this is a real concern.

A second concern at this stage of sampling was the exclusion of women “who had only ever had non-medical terminations”. It’s not clear what this exclusion refers to and whether they got a stage two questionnaire and were subsequently excluded or did not receive a stage two questionnaire? Is this women who reported positive home pregnancy tests but no clinical pregnancy? Why are they excluded?

In the third stage, women with a pregnancy loss that was recent (since 1995) but not their last reported pregnancy were asked for more information about the pregnancy loss. Why was this time restriction imposed? Minimally, the impact of adding this subset to the relation between the sample and the population should be addressed.

Using the stage two data, the authors also report using women as both cases and controls in a case-control study of miscarriage. The most recent (control) pregnancy should be excluded from comparisons that involve the third stage sample or more complex statistical methods are needed to address the lack of independence of outcomes to the same woman.

The definition of multiple delivery is also non-standard and it is not clear why the authors would use it. Multiple gestation is an important reproductive event and it appears poorly recorded in this dataset. Beyond selective reductions (which are themselves uncommon), the likelihood of a twin pregnancy with an identified early loss of one twin is also low. What is the authors rationale in calling that a singleton birth?

Are the data sound and well controlled?

The data that were actually collected may be sound but readers would have more confidence if the quality assurance procedures for data handling were mentioned and if the authors examined the reliability of data provided by women more than once (for example comparing similar items, if any exist, from stage one, two or three questionnaires). The notion that this data can represent the underlying population is overstated due to the concerns raised above.

For comparisons to the national datasets, the authors should consider restricting their study sample to only married women. That would allow a direct comparison between the population and their sample for at least that subgroup of women.

Does the manuscript adhere to the relevant standards for reporting and data deposition?

The authors do not report whether or not the study was conducted under the approval of a review board or address other considerations such as protection of privacy and confidentiality of data storage.

The clarity of reporting could be improved. There are a few errors in the numbers of subgroups and it is extremely hard to follow the sampling stages as written in the text. I think a figure or flow diagram that shows the origin of the sample at each stage and the number of respondents (n and percent) would be most helpful. For example, we do not know the number of women in the electoral role or the number of women selected in the initial sample. We know there are 60,814 women selected after application of the name probability selection, but it is not clear how the authors get to 10,828 stage two questionnaires? Etc..

Are the discussion and conclusions well balanced and adequately supported by the data?
No, I'm afraid not. The authors misuse the concept of random sample and overstate the generalizability of the study group(s). The data may still be of great value, but the focus should be on internal validity rather than any attempt to claim external validity (unless the authors are prepared to address the concerns about sampling raised above).

Do the title and abstract accurately convey what has been found?

As a descriptive paper, this is fairly straightforward. However, the abstract does not convey the complexity of the sampling and the data presented are not sufficient to say that the sample is representative of the general population.

Is the writing acceptable?

The paper is generally well written but some of the descriptions of the methods could be clarified as noted above.

Major compulsory revisions

Focus on the potential strengths of the internal validity of the data and drop the repeated claims of generalizability to the population. The paper needs to be refocused and the limitations of the non-standard sampling and definitions should be addressed directly. If the authors can conduct the additional analyses mentioned above (if the data are available, etc..) they may be able to retrieve a claim for generalizability but it seems unlikely from the data presented.

Next step?

Unable to decide on acceptance or rejection. The problems with the design are quite overwhelming especially given the authors desire to use the data to estimate population prevalences.

What next?: Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable

Statistical review: No

Declaration of competing interests:

None