

Reviewer's report

Title: Epithelial cells and abnormal cytology in nipple aspirate fluid and subsequent breast cancer risk: a historic prospective study

Version: 2 Date: 12 August 2007

Reviewer: Gertrude C Buehring

Reviewer's report:

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

1. p. 2, last 3 lines, p. 5, lines 15-16: The description of the two subgroups, women with questionnaire data, and those with NAF visits beginning 1988, is very confusing. These two groups do not seem to have been separated in two groups for statistical analysis, so calling them two subgroups seems inappropriate. Do authors just mean that there were two different ways of obtaining information about whether the women had breast cancer, one way from the questionnaire and one from the cancer registry? If this is the case, it would seem clearer to omit the last 3 lines of p. 2 and then explain on p. 4 as the second paragraph under "follow-up protocol" how data on breast cancer status was obtained from the cancer registry for subjects who did not return a questionnaire.

Related to this issue is the text in the last paragraph of p. 3 carrying over to p. 4. The sentence "The final model...is composed of women who completed questionnaire data and/or women who were seen by D. Sartorius..." Weren't all the subjects seen by Dr. Sartorius? It would be less confusing if the content of this paragraph were incorporated into the paragraph suggested above under follow-up protocol.

p.5, "Of the breast cancer cases reported, 75 were self-reported [only] and the remaining 18 were confirmed via the CCR." Were any from the CCR only?

2. p. 4, third line from bottom, p. 5, line3, and several other places in the manuscript: The term "insufficient specimen" is not defined. I assume it means a sample of fluid with no epithelial cells, but how can I be sure it doesn't mean insufficient fluid to process the sample? It isn't defined anywhere. It would be clearer to say "fluid without epithelial cells."

3. p. 4, last line: "For women with multiple NAF results, the most severe result was used." Doesn't this bias the data? For women who were sampled only once, and perhaps had no fluid or fluid with no cells, how do you know that they might not have had a sample with epithelial cells on a subsequent sampling? If the authors have the data for the first sampling of all subjects, it would be important to determine if the results would come out differently if only first-try data were used.

4. p. 5, line 10: Although both models were adjusted for age at the time of the NAF visit, it would be interesting to see if stratification into age groups that roughly represented pre- and post-menopausal groups would show different ORs. This would be important because pre- and post-menopausal breast cancer seem to have such different characteristics. Could this be done?

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

1. p. 3, line 12: What covariate information other than age was collected? Was any of it analyzed? If not, why not?

2. p. 7, line 13: If references 7 and 9 are review articles, perhaps it would be better to say "(reviewed 7,9). If they are primary articles, then other primary articles published in the 1970s should also be cited.

3. References: There are no page numbers on references 5 and 7.

4. Table 1: the word "results" is somewhat ambiguous. It would be clearer to use "data" instead.

5. Table 2. Indicate whether the differences between "not in model" and "In model" are statistically significant.

Discretionary Revisions (which the author can choose to ignore)

1. p. 3, line 8: Would sound better as "Subjects were a cohort of 3,203..."

2. p. 3, line 14: "...by a team of registered nurses in Santa Barbara, California." would sound better.

3. The title might sound better as "Epithelial cells in nipple aspirate fluid ..." since abnormal cytology is part of the epithelial cells.

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

Level of interest: An article of importance in its field

Quality of written English: Acceptable

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:

I declare that I have no competing interests.