Reviewer's report

Title: Proteins and carbohydrates in nipple aspirate fluid predict the presence of atypia and cancer in women requiring diagnostic breast biopsy

Version: 2 Date: 27 September 2011

Reviewer: Gertrude C Buehring

Reviewer's report:

Major compulsory revisions:

Overall, the manuscript is well-written and the data interesting. However, there are a few issues indicated below. Hopefully, these comments will be helpful in improving the manuscript.

1. I find the use of ND (nipple discharge) as the general term for all specimens somewhat confusing, since NAF (nipple aspirate fluid) has to be suctioned out and is therefore not what most people are used to calling a discharge. Wouldn't NF (nipple fluid) be a better general overall term? The two subdivisions could then be NAF (aspirated fluid) and ND (spontaneous nipple discharge). What is “pathologic nipple discharge”? This term suddenly appears on page 9, but is not defined on the bottom of page 4 along with the other terms for nipple fluid. What makes PND pathologic and how are the authors distinguishing it from spontaneous discharge? The authors should define their categories of nipple fluid more clearly.

2. There is also some confusion in the pathology terminology. I suggest Fitzgibbons et al. Arch Pathol Lab Med 122: 1053-1055, 1998 as the accepted reference for breast pathology terminology (Consensus statement of the College of American Pathologists). You might consider having three classifications: normal = all classifications that are not considered to carry an increased risk of developing into cancer (fibrocystic disease, fibroadenoma, hyperplasia, and other pathologies listed in Fitzgibbons et al.); precancerous = classifications with an increased risk of developing into cancer (atypical hyperplasia, solitary papilloma); and cancer (DCIS and invasive cancer). Many pathologists do not consider DCIS as cancer but rather precancerous. Whatever you decide, define the categories and analyze the data accordingly. Your definitions on p. 17-18 are fairly clear, but these should be presented earlier in the methods sections so that the reader can understand the grouping discussed in the statistics section. However, on p. 17-18 it is unclear what “family history” is doing in the pathology definition. Do all of the subjects have a family history? Family history of only one first-degree relative is not considered high risk. Please explain the role of family history in this study.

3. It is worrisome that there is no method used to normalize the concentration of the nipple fluids. The enormity of the volume range suggests that some specimens with larger volumes could be more dilute. Did the authors test to see if
there is an inverse relationship between sample volume and the concentrations of the proteins they are measuring and between sample volume and age? Previous studies on NAF showed that younger subjects had volumes of NAF larger than older subjects. Could the target protein concentrations be normalized to some housekeeping protein or total protein? I know this can be a difficult issue. However, in the clinical laboratory the concentration of serum albumin and globulin is used as a yardstick against which to measure concentrations of other serum substances. For instance, if the albumin is elevated above the normal range, it is a sign that the individual is dehydrated and if values for other blood components are elevated, they could simply be due to dehydration and not disease.

4. Page 8, line 1: Please indicate what tests were used to prove that the distribution after log transformation was normal and please confirm that the distribution passed such a test.

5. Could the authors say a little more about the potential usefulness clinically, viz. what percentage of pre- and postmenopausal women would yield a suitable specimen that could be analyzed?

Minor essential revisions

1. The acronyms AUC and ROC should be defined when first used on p. 8.

2. Page 8, line 6: HRT is not defined. How does it differ from “hormone replacement”?

3. Final paragraph: “Each” is repeated in the first line.

**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.