Reviewer’s report

Title: Breast cancer epidemiology according to recognized breast cancer risk factors in the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial Cohort.

Version: 1 Date: 5 July 2008

Reviewer: Tina Clarke

Reviewer’s report:

This is a well-organized manuscript addressing a clearly defined research question: do the "established risk factors for breast cancer" as identified in myriad small populations in prior years and as confirmed by the Gail model predict breast cancer risk in a large screening cohort? This data resource is acknowledged to have been collected for other analytic purposes but large enough to be relevant to breast cancer (when treated as a single entity); the analytic methods employed are appropriate and clearly documented according to standards for epidemiologic studies. The title and abstract are accurate descriptions of the work and its findings. The interpretations and conclusions are adequately supported by the data, prior literature appropriately referenced (with minor exceptions described below). Its major limitation as I see it includes the inherent lack of several confounding factors in the data resource that could modify magnitude of hazard ratios reported and inadequate discussion of the impact of this limitation on study results.

DISCRETIONARY REVISIONS

Abstract/introduction sections: these do not provide enough discussion of the relevance of the research question for risk prediction (e.g. the Gail model). This relevance is discussed in the discussion section but also belongs in the abstract and introduction.

Methods section: could benefit from some more detail regarding questionnaire content and detail as relevant to risk factors included in analysis. For example, was menopausal hormone therapy use ascertained with detail regarding formulation?

Results section: first three sentences of text could stand to have percentages included in addition to absolute numbers to provide a better sense of the prevalence of these characteristics in this population...these are available in the table

MINOR ESSENTIAL REVISIONS

Citations in discussion section: The sentence describing risk factors not addressed by the survey questionnaire cites one paper documenting associations with breast feeding but not alcohol consumption (could use Beral et
al. meta-analysis here as well) or physical activity. These references should be added.

MAJOR COMPULSORY REVISIONS

Uncontrolled confounders in this analysis: It is mentioned in the methods section that the PLCO trial initially excluded women with bilateral oophorectomy or who were taking tamoxifen but later allowed these women to enroll. Were these women included in the present analysis? Were these characteristics adjusted for in the present regressions? The authors should clarify if these women were excluded and if not, the characteristics added to the tables as established risk factors for breast cancer and appropriate hazard ratios calculated. It is also mentioned in the discussion section that the study questionnaire did NOT include information regarding alcohol consumption, physical activity, or lactation history but there is no critical commentary provided to understand how the exclusion of these factors might have influenced the reported hazard ratios. Lastly, menopausal hormone therapy is not separated in these analyses by progestin-containing formulation, which is relevant because only estrogen/progestin (EP)-containing therapies are definitively associated with risk at this time. It is unclear from the methods whether grouping together all hormone therapies represents a analytic choice of the authors or not. If possible, the data should be reanalyzed to separate EP use from other forms of use.

Discussion of risk prediction for breast cancer as a single entity: Breast cancer is increasingly recognized as a group of etiologically heterogeneous subtypes. I am unsure how the cutting edge risk prediction work is incorporating this essential observation into future models (for example, are new models being developed to predict estrogen-sensitive breast cancers only?). The discussion section does not address breast cancer etiologic heterogeneity and its implications for the study findings or relevance. It should do so.

Discussion of study population selection bias as explanation for findings: Table 1 shows distributions of many demographic factors in this study population...do these distributions differ dramatically from the prior studies which detected different associations with the lifestyle risk factors that are likely associated with participation in this cohort? Like many contemporary research endeavors, it would be reasonable that this cohort overrepresents white and college-educated women, who may be more likely to agree to participate in research. Wouldn't a corresponding reduction in the distribution of several key risk factors also be an explanation for some of the attenuated hazard ratios observed? This possibility should be discussed as an alternative interpretation of the reasons for the changing risk factor profiles suggested by this study.

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

Quality of written English: Acceptable
**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I have served as an expert witness for breast cancer patient plaintiffs in ongoing hormone therapy litigation.