Reviewer's report

Title: Inter-observer agreement according to three methods of evaluating mammographic density and parenchymal pattern in a case control study: impact on risk prediction

Version: 1 Date: 14 December 2014

Reviewer: Jingmei Li

Reviewer's report:

The authors compare inter-observer agreement and differences among three different methods of characterizing mammographic density - BIRADS, TABAR and area-based percent mammographic density. In addition to comparing risk by different categories, often by increasing density, the authors also combined categories into low/high risk, which decreased inter-observer variability. The conclusion is that the three methods generally predict breast cancer risk, but are associated with different effect sizes. The authors also highlighted that the Tabar patterns are associated with the highest discrimination of breast cancer risk in terms of odds ratios.

The question posed by the authors was well-defined, and the methods are extremely detailed and well-described. Although the parent screening study is huge (n=14736), the number of cases used for the breast cancer case-control study is relatively small (n=122).

Discretionary Revisions

1. Scatter plot and Blandt Altman plot: the numbers on the axes have commas instead of "." to denote decimal points.

2. Do the results change if the DCIS cases were removed? Discuss the potential effect the DCIS cases would have on the effect sizes of breast cancer risk.

3. p14 lines 14 and 17. The result is either significant or it isn't. If the p-value is higher than 0.05, it should be called 'non-significant' and written up as such.

4. Apart from BMI, menopausal status is also strongly associated with breast density. The lack of information on menopausal status should be acknowledged too.

5. How does the risk prediction results look if you average the scores by both readers and use that in the model?

6. p20 line 17. What is the overlap of women who are considered high risk among the three different methods? Are the high risk women always the same women, i.e. what is the proportion of women who are consistently judged to be in a high risk category?
Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Needs some language corrections before being published

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.