Reviewer’s report

Title: Breast cancer stage at diagnosis and area-based socioeconomic status: A multicenter 10-year retrospective clinical epidemiological study in China

Version: 2 Date: 12 January 2012

Reviewer: Sandi Pruitt

Reviewer’s report:

Review of Biomed Central : BMC Cancer

This manuscript examines the role of area SES on breast cancer in China and represents, to my knowledge, the first attempt to explore this topic in a developing country. Overall, the data and analysis are strong and the hypothesis is worthwhile. Overall, the writing is very good. A few rather minor edits for clarity and standard English could benefit the readability, however.

Major Compulsory Revisions:

1. My primary concern with this paper is related to the construct of “area-based socioeconomic status” here. The authors measured indicators of SES across 7 very large geographic regions and do not control for other differences between these areas. For example, the differences across these areas could be a result of other regional differences in culture, behaviors, language, health care access, health care quality, ethnicity, age-distribution, or norms. In the U.S. research has shown that area SES is most reliably and validly measured when measured at a very small geographic scale, not at a large geographic scale as these authors measure it. It would strengthen the paper to discuss the validity of this approach and the conceptual meaning of area SES in China. It may be useful to control for other area-level covariates in the analysis. The authors may also want to consider discussing in the limitations the fact that instead of measuring area SES, they may be simply identifying geographic variation.

2. Methods-Please provide more detail what you mean by “cluster analysis.” Do you mean principal components analysis? More information should be provided by what this means and how it was conducted in the methods section.

3. Methods-Please provide more information on how the categories of SEIs and SES were categorized and measured as “highest, high, and low.” What are the mean scores and range for these categories? How were these created?

4. Methods-The rationale for stratifying the analysis by area SES is not clear. Please provide a rationale for why you stratified by this variable. Results-Table 3: Why do you only present stratified analyses here? Why don’t you present the total analysis as well? It is common practice to test whether area SES is associated with an outcome after controlling for individual SES? Please explain why the analysis wasn’t done this way.

5. Methods- Provide more information regarding how all the variables were
measured and categorized. Were they all taken verbatim from the medical record? Were they re-categorized or changed for this analysis? For example, some of my questions about the variables are: is cancer stage determined by a pathologist? Is employment status as measured in your analysis really captured in the electronic medical record (manual, homemaker, etc)? Is this based on self-report of the patient or from some kind of government record? Is this a reliable measure?

6. Results- If you create and use a SES scale, it would be helpful to describe the scale’s psychometric properties. At the very least, provide the scale’s Cronbach’s alpha. Also describe if the SEI measures were reverse-coded in the creation of the scale, the scale’s scoring range, and it’s mean.

Discretionary Revisions
1. In the U.S., urban/rural status is not highly correlated with socioeconomic status and is generally considered a different construct separate from SES. Please provide a brief rationale why you include urban/rural status as a marker of SES. I assume because it is highly correlated with SES in China.

2. It would help to clarify which categories of “highest” “high” and “low” variables in Table 2 represent “higher” SES. For example, is the ratio of FPI reverse-coded?

3. Why are there 4 groups (highest, higher, high, low) for the PU/PR ratio instead of 3 as used for the other SEIs?

Minor Essential Revisions
1. The terminology “composited” and “incorporated” and “composited categorized” used to describe SEIs and SES measures are not standard terminology. It appears you mean “composite” SES “scale.” I also do not know what you mean by “incorporated SEIs.”

2. The terminology “adjacent” stage is not standard. It may be clearer to substitute the terminology “later” or “more advanced” stage instead.

3. The term “ANOVA-SNK” (p.5) is unclear. What is SNK? Also, why aren’t the results of the ANNOVA provided?

4. The final statement in the conclusion is somewhat awkward.

5. Is there no data available at all regarding mammography or breast cancer screening? I’m just curious because this would be very helpful to your manuscript to know about history of screening.

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’