Reviewer's report

Title: Marriage, interpregnancy interval, neighborhood, and small for gestational age birth: a retrospective cohort study

Version: 1 Date: 14 June 2007

Reviewer: Bao-Ping Zhu

Reviewer's report:

General

This study explored the interactions of three well-known risk factors for SGA birth – marital status, interpregnancy interval, and neighborhood characteristics. The findings of this paper shed some additional light on the complex inter-relationships among these factors in relation to SGA birth, which should serve as a basis for further research on this issue.

-----------------------------------------------------------------------------------------------------

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

My biggest concern is the way interpregnancy interval was categorized. Based on their descriptions in the Methods section, the authors decided on the three categories based on sample size alone, whereas it is more appropriate to decided on the categories based on the actual association observed between interpregnancy interval and SGA birth, or on the accepted categories in previously published articles (such as <18, 18-23, >23 months, as was used in a recent meta-analysis published in JAMA [ref 1 in the manuscript]). The way the categories are defined in the current version of the paper renders it difficult to compare the findings of this paper with other published works.

-----------------------------------------------------------------------------------------------------

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

Consider changing “birth categories” into “interpregnancy interval categories” throughout the paper. I understand that “first-born” is not an IPI category per se; however, it can be considered as a special case. To me, “birth category” is vague, uninformative, even misleading.

Title: Consider changing it to: “The joint effect of marital status, interpregnancy interval, and neighborhood characteristics on small-for-gestational-age birth,” to better reflect the content of the paper.

Abstract: Consider changing the last sentence in “Methods” to: “Multilevel logistic regression was employed to elucidate the associations.”

p. 6., 2nd paragraph (under “Data”): I wonder how gestational age is determined on the Canadian birth certificate? Is it LMP-based? If yes, how was the missing values dealt with? It would be helpful for readers to know these details. In the same paragraph, did the authors use years of education as a continuous variable in the logistic regression? If yes, did the authors investigate whether it was appropriate to do so - i.e., whether the relationship between years of education and SGA was log-linear? If it is not log-linear, the authors will need to categorize education, based on the shape of the curve.

p. 7, 3rd paragraph (under “Exclusion criteria and missing data”): The way the education variable was imputed could cause misclassification. There are more sophisticated and better imputation methods available. However, I wonder whether the authors had considered simply creating a dummy variable representing those with missing education, and if yes, whether the results were any different?

p. 8, 2nd paragraph. Since the authors pooled many years of data, there are likely biological siblings (i.e., infants born to the same mother) in the dataset, whose data are likely correlated. The authors should consider linking these infants to their biological mothers, and consider another level of intra-class correlation (in addition to neighborhood dependency). If this is impossible, they should acknowledge this as a limitation in the discussion section.
As I said earlier, I have a big problem with the way interpregnancy interval was categorized.

p.10, 1st paragraph, the sentence that starts with “For subsequent born infants…” is not informative. These %s are strictly determined by the cut-points the authors had chosen.

p.10, 2nd paragraph. The X2 and p-values are not very informative. They are determined by the sample size. A more appropriate way to present the data is to describe the magnitudes of the associations.

p.10, 3rd paragraph. I have a problem with the statement in the 1st sentence of the paragraph. The lack of statistical significance could be simply due to the smaller sample size. I would delete this statement.

p.13, last paragraph. Ref. 40 is not related to the current topic under discussion. (The authors might want to investigate the other references more completely on their relevance to their study.)

p.14, 2nd paragraph. I suspect a major reason that the attributable fraction was different between this paper and ref. 34 is that IPI was defined differently.

p.14, 3rd paragraph. I wouldn’t characterize the lack of information about years of residency as “unavoidable” unless the data are not collected on the Canadian birth certificate.

References: Check the references to ensure uniformity (e.g., ref. 27).

Tables and graphs: All captions should contain information about what, who, where, when.

Figures 1-2. The vertical axis should be on log-scale. Otherwise the graph will give misleading impressions.

Discretionary Revisions (which the author can choose to ignore)

I was curious as to why the authors chose to investigate these three variables, not the other important risk factors, such as education, income, and employment?

What next?: Accept after minor essential revisions

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

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:

I declare that I have no competing interests.