Reviewer's report

Title: A flexible Bayesian hierarchical model of preterm birth risk among US Hispanic subgroups in relation to maternal nativity and education

Version: 2 Date: 19 November 2010

Reviewer: Nathalie Auger

Reviewer's report:

Dear Editor,

Thank you for asking me to review this well-written manuscript addressing the relation between education, nativity, Hispanic ethnicity and preterm birth. This review is written from the perspective of someone who is not a specialist in Bayesian analyses, but has basic familiarity with this statistical method.

Abstract

1) Last sentence – it’s not clear from the abstract how the last statement was deduced from the previous elements in the abstract. I gather the last sentence refers to the Hispanic Paradox, but would it not be more helpful to state that the monotonic influence of education observed in previous studies does not reflect the influence of the spectrum of education, especially across different ethnicities? Or that greater attention to Puerto Rican women including preterm birth prevention programs may be warranted?

Background

2) Paragraph 1 – In general, I think it would be helpful to reformulate the first two sentences to make it easier for the reader to follow the arguments. It’s not completely clear how acculturation ties in. It would also be helpful to clarify the definition of “Hispanic Paradox”, in particular the comparison group. That is, who are “other disadvantaged groups” – does it refer to Blacks that are undereducated or to some other group? My own personal opinion is that this paragraph could start off with the third sentence, and incorporate the material in the first two sentences later on.

3) Last paragraph. Consider mentioning that preterm birth may be more relevant from a public health perspective than low birth weight. Recent studies have demonstrated the importance of prematurity on morbidity and mortality throughout life (e.g., Saigal S, Doyle LW. An overview of mortality and sequelae of preterm birth from infancy to adulthood. Lancet 2008; 371(9608):261-269 or Moster D, Lie RT, Markestad T. Long-term medical and social consequences of preterm birth. N Engl J Med 2008; 359(3):262-273), while the influence of low birth weight on mortality is less clear. Furthermore, rates of preterm birth are increasing in most countries (but decreasing for low birth weight in some places, e.g. Canada).

4) It is always helpful to have a sentence stating the objective explicitly. It took
me a while to understand that the underlying objective was to shed light on previous research that used simpler methodologies that may have led to incomplete pictures of the relation between education, nativity, ethnicity and birth outcomes (and that Hispanic Paradox per se was not the main focus).

**Methods**

5) Why was the clinical estimate used instead of date of last menstrual period?

6) Please clarify the education variable as the sentence seems to suggest that data were available after 17 years, but not used. Is this correct? Also, what does “using a flexible regression model” mean?

7) Could you provide a reference for the covariates, especially for the Kessner Index.

8) Statistical analysis- It would be helpful for readers not familiar with Bayesian analyses to clarify terms such as “hyper-parameters” or “borrow strength”, but this may not be necessary since I understand that the paper probably targets a specific population. However, there are a few things that would benefit from clarification, including the definition of “m” and “x” in equation 1, and a more clear justification of why 8, 11, and 13 were selected for the knot locations. Do these cut-points have special significance in the US? Is a high school diploma obtained at 11 years? There also were many assumptions and I feel the article would benefit from making them more explicit or providing justification, and perhaps discussing the implications. For instance, the precision terms for some of the alpha coefficients were forced to be identical – why? Was it to make the models easier to handle? If so, what are the implications?

**Results/Tables/Figures**

9) Paragraph 2, sentence 2. The table suggests South America is not an exception.

10) Table 1 – the variables listed in the table seem different than those listed in the methods (and some variables listed in the methods do not appear in the table). Is there a reason?

11) I had a hard time with figures 1 and 2 – the colour did not illustrate the differences well on my printout. But this forced me to look a bit more closely at Appendix A1-A4, which were significantly clearer and my first thought was that those figures ought to be used instead (and that figures 1 and 2 were not necessarily needed). But this choice is up to the authors.

12) Risk differences were calculated for nativity, but could have been calculated for ethnicity. This is all right, but think it would be helpful for the reader to understand why (perhaps by more clearly stating in the objective that special emphasis was placed on differences between native and foreign born groups).

**Discussion**

13) Would benefit from mentioning the limits of clinical estimate of gestational age
Minor

Introduction - Paragraph 1 - change if to of (line 6), and insert a period for the last sentence

Thank you for asking me to read this interesting contribution that hopefully will lead to greater application of Bayesian methods in perinatal epidemiology.

**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 declare that I have no competing interests