Referee #?

PONE-D-19-26817

Comments on "The role of neighborhood socioeconomic status in large for gestational age: a nationwide study"

1. <u>Overall evaluation</u>

This study evaluates the correlation between zip-code level neighborhood socioeconomic status (SES) and the risk of large for gestational age (LGA), focusing on newborn hospital birth admission claims in an urban area – Marseille – extracted from nationally representative administrative data called the *French National Uniform Hospital Discharge Database*. Statistical strategy is straightforward, applying multivariate logistic regression model. The result shows that lower zip-code level of neighborhood SES tend to be associated with LGA significantly, after adjusting for mother's age, diabetes, obesity, smoking behavior and multiparity. Further, supposed that mother's age is a proxy of duration of residency, the author(s) discuss that older mothers living in the area with lower SES might experience a cumulative disease risk over time and therefore, the intensification of the correlation between neighborhood SES and LGA increases along with mother's age.

I fully agreed with the contributions and novelties highlighted by the author(s), such that this study sheds a light on the issue under the universal maternity coverage in a developed country – France – other than Northern American countries; and also it could provide clinicians additional information to help them understand mothers' risks of LGA births, along with conventional ones, such as maternal diabetes, obesity and age. However, I think that the author(s) had better to explain carefully about the situation in France as a background to persuade international readers. I would like to provide both major and minor comments as follows. Hopefully, they are helpful for the author(s) to improve this paper.

2. <u>Major Comments</u>
2-1.   <u>Title</u>

The current title "The role of neighborhood socioeconomic status in large for gestational age: a nationwide study" is unclear about target population to be analyzed. Further, although this study focused only on an urban area in an urban area – Marseille –, the subtitle ("a nationwide study") might mislead the readers. So, I strongly suggest the author(s) to revised the title to the one which reflects the contents of this study more precisely.

2-2. **More detailed explanation may be helpful for international readers to understand the situation in France as a background information**

In the section of "BACKGROUND", the author(s) had better to provide more detailed explanation how serious LGA births has been becoming, not only in France but also in other developed countries, showing some basic statistics (time-trend LGA births, if possible).

Also, the author(s) emphasizes high standards of maternity care in France due to guaranteed access to care regardless of costs for pregnant women in 14th line of page 3, without any reference. So, the author(s) need to show some reference for this. Then, the author(s) had better to show basic statistics and/or to conduct institutional explanation for proving how high the standards of maternity care in France are, compared to other developed countries like OECD countries.

Adding such more detailed clarifications may help to persuade international readers how important this study is.

2-3. **Why Marseille?**

As long as I noticed, in the section of "METHODOLOGY", there are no explanations why the author(s) chose Marseille from a nationwide population-based database in France. Then, from the 2nd bottom line in page 9 through page 10, the author(s) briefly mentioned why they chose an urban large city for this study, such that "Lastly, the accuracy of geographical methods based on residential postal code can nevertheless be put into question in studies including urban and rural populations with nonhomogeneous social neighborhood characteristics. This study avoided such a bias by limiting investigations to a single large city". The author(s) definitely should reallocate this sentence to the "METHODOLOGY" section. Further, the author(s) should clarify the reason why they chose Marseille among multiple urban areas/cities in France and they have to discuss the possible *sampling bias issues* on the estimation caused by focusing on a large city like Marseille in "DISCUSSION" section.

2-4. **Statistical analysis & results**

2-4-1. A flowchart for the procedure of extracting the data to be analyzed is necessary

In the first paragraph of "RESULTS" section in page 6, the author(s) describe how to extract the data to be analyzed in this study. I suggest to bring this explanation to "METHODOLOGY" section and show it by a figure of flowchart. It may be very

helpful for the readers to understand how the data is constructed.

2-4-2. Justification for the threshold "p-values<0.2"

In the 7-9th line of page 5, the author(s) said "Variables relevant to the model were selected based on a threshold p<0.2". Please provide the rationale for choosing this threshold. Otherwise, the selection of explanatory variables seems to be with some 'intention'. If the author(s) do not have any clear justifications for this, they should to some stepwise logistic regression analyses.

2-4-3. Sensitivity analysis excluding/imputing missing data on maternal age

From the bottom line of page 5 through page 6, the author(s) mentioned the above sensitivity analyses. However, I do not see any results for this. Please show the results of sensitivity analysis, at least in supplementary tables. Also, if the author(s) did some imputation for missing maternal age, they should briefly explain which imputation method they used, either in the main body of text or in the appendix.

2-4-4. Test statistics for regression analyses are necessary to be shown in Table 3

Test statistics for validity of the model should be shown, such as log likelihood, pseudo-R2 values, and Akaike's Information Criterion (AIC).

2-4-5. Univariate results are not necessary to be shown in Table 3

In Table 3, the author(s) show both results of univariate and multivariate logistic regression analyses. However, I do not see any meaning to show univariate results (besides, the results do not seem to be different much). Rather, it is more interesting to show the results of supplementary tables 2 and 3. Besides, in page 7, the author(s) explain the results (ORs) of these supplementary tables in detail. For example, the author(s) said "The strength of the association between SES and LGA increased with maternal age, reaching an aOR of 1.59 (95% CI: 1.26; 1.78 p<0.0001)...", which cannot be seen in any tables in the main body of manuscript.

2-4-6. How to estimate the results for Figure 1 & supplementary table 3

I do not see how to extract the results of Figure 1 (supplementary table 3). I guess the author(s) introduce interactive terms of maternal age and NDI levels into logistic regression analysis. If possible, please clarify the estimation model in "METHODOLOGY" section. Otherwise, the reader must be lost in a series of explanation about the results of Figure 1 in page 7.

3. Minor Comments

(1) In page 4, the translation of the name of dataset sounds a little strange. How about "the French National Uniform Hospital Discharge Data" or "the French National Uniform Hospital Discharge Database". I do not think that the author(s) need to repeat "Data Set Database".

(2) In the 8th line of page 5, what is (Audipog) (19)? Please clarify.

(3) In "RESULTS" in page 6, the author(s) said "The rates of infants born to mothers with advanced age (more than 35 years)" is "51%", which is different from the value ("20.1% for age 35 and older") shown in Table 1.

(4) Same as (1), In "RESULTS-LGA population characteristics", the author(s) said "In comparison with AGA infants, LGA infants were more likely to be born to older (maternal age more than 35 years) and multiparous mothers (p<0.0001)". However, Table 1 shows an opposite result, such that the ratio of Cesarean delivery tends to be significantly larger for LGA infants so that LGA infants were less likely to be born to multiparous mothers.

(5) At the second line from the bottom of page 6, the author(s) explains about GA (p=0.81). Then, again, the author mentioned about GA at the first line of page 7, such as "While the mean GA and preterm birth rate were comparable". Better to erase either one.

(6) At the forth line in page 7, the author(s) mentioned "with higher rates of hypoglycemia, hypocalcemia". Yet, Table 1 does not show the rates of these variables. Please add these in Table 1, if the author(s) use these variables for the further statistical analyses.

(7) In "RESULTS-Neighborhood deprivation and perinatal outcomes" in page 7, the author(s) mentioned "… and to be prenatally exposed to maternal smoking", which is different from the results shown in Table 2. Table 2 shows that infants born to mothers living in Q2 and Q3 of NDI are more likely to expose to maternal smoking (1.9%), than Q4 (1.4%) and Q1 (0.8%). Please revise the sentence.

(8) Again, in the same section as above (5), the author(s) repeatedly explain about the ratio of LGA. In the 2-3 lines of this section, the author(s) said "The proportion of LGA varies by NDI levels, reaching 12% in NDI level 4". Then, in the 3-4th lines from the bottom of this section, the author(s) repeatedly explain "…, while we observe a higher rate of LGA with the NDI (12.3% in NDI level 4 vs 9.9% in NDI level 1, p<0.0001)". Please drop the first part.

(9) As mentioned in my major comments, I do not see any reasons why the author(s) show "Univariate analysis" in Table 3, which is not necessary. Instead, the author(s) should show the results of supplemental table 2 & supplemental data table 3.