Author’s response to reviews

Title: Using ecological socioeconomic position (SEP) measures to deal with sample bias introduced by incomplete individual-level measures: Inequalities in breast cancer stage at diagnosis as an example

Authors:

Sébastien Lamy (sebastien.lamy@inserm.fr)
Molinié Florence (florence.molinie@chu-nantes.fr)
Daubisse-Marliac Laetitia (laetitia.daubisse-marliac@inserm.fr)
Cowppli-Bony Anne (anne.cowpplibony@chu-nantes.fr)
Ayrault-Piault Stéphanie (stephanie.ayrault@chu-nantes.fr)
Fournier Evelyne (Evelyne.Fournier@unige.ch)
Woronoff Anne-Sophie (asworonoff@chu-besancon.fr)
Delpierre Cyrille (cyrille.delpierre@inserm.fr)
Grosclaude Pascale (pascale.grosclaude@inserm.fr)

Version: 1 Date: 25 May 2019

Author’s response to reviews:

Responses to the reviewers

Title: Using ecological socioeconomic position (SEP) measures to deal with sample bias introduced by incomplete individual-level measures: Inequalities in breast cancer stage at diagnosis as an example

We would like to thank both the editors and the reviewers for their helpful suggestions to strengthen our manuscript. We respond hereafter to their remarks, and the manuscript has been modified in accordance with their suggestions. The reviewers’ comments are in bold, the authors’ responses are in italic below, and samples of the manuscript are underlined.
Reviewer reports:

Marek Brabec (Reviewer 1):

This is a study of substantial practical interest. It shows and demonstrates clearly the dangers of relying on the complete-case analysis only. It uses relatively simple statistical tools and its message is clear. It is well suited for the BMC Public Health readership.

The paper is ready to be published in its present form.

The Authors’ response: We thank the Reviewer 1 for his comment.

Tzu-An Chen (Reviewer 2):

This manuscript was generally well-written. This study examined the effect of missing individual-level SEP and using area-level SEP on the association with breast cancer stage at diagnosis. I have several concerns that temper my enthusiasm for this work in its current form. The authors are commended for their work and should consider the following points to help clarify and potential strengthen the impact of these data.

1. The background of the influence of SEP on cancer (or breast cancer) should be strengthened, rather than a single sentence, as it was the major association that the authors intended to explore. Currently, there was only one sentence describing SEP and cancer in entire background section.

The authors’ response: The paper deals with sample bias introduced by the use of incomplete individual-level measures in the study of social inequalities in health. We based our discussion on the study of the relationship linking socioeconomic position (SEP) on breast cancer stage at diagnosis. Although that the patient’s SEP-cancer stage at diagnosis relationship only serves as example, we agree that the paper would benefit from more precision. We added the sense of the SEP-cancer stage at diagnosis relationship supported by the literature at the end of the introduction section (lines 86 to 89, page 4)

2. The authors should examine the missing pattern of the data: Missing complete at random, missing at random, or missing not at random. Sometimes analyses of complete cases will not lead to bias.

The authors’ response:

We agree with Reviewer 2 that sometimes analyses of complete cases will not lead to bias, especially when data missed according to a MCAR pattern. As we wrote in the initial version of
the manuscript, in the fourth paragraph of the discussion section (lines 277 to 279, page 15), from the results of the “missing data analysis” we may reasonably think that the missing data follow a MAR pattern. Therefore, the use of multiple imputation seems not irrelevant.

3. Please define these acronyms "TNM", "HER2" and "SBR".

The authors’ response: We added the acronyms definitions in the text at line 109 for TNM, and lines 141 and 142 for HER2 and SBR.

4. I appreciate the authors including Figure 1 within the paper.

The authors’ response: nothing to add.

5. Statistical analysis: please indicate what analysis was done to compare between respondents vs. non-respondents in the statistical analysis section even though it was described in Table 1 footnote.

The authors’ response: This was added in line 149.

6. Statistical analysis: Please indicate what analyses were conducted for these configurations listed. Please also include the designated significant level.

The authors’ response: This was added in lines 173 and 175.

7. Table 1. Please keep 1 or 2 decimals for %.

The authors’ response: This was done.

8. If EDI was not significant between respondents and non-respondents, why did the authors want to use that in the imputation? The authors should provide stronger support and maybe move the spearman correlation in Discussion to earlier section.

The authors’ response: We added some sensitivity results that supports our choice at the end of the second paragraph of the result section. However, we kept the discussion about the spearman correlation in the fourth paragraph of the discussion section, but it was modified as follow:
“Although results showed differences in the proportion of patients living in the areas with the highest deprivation level between respondents and non-respondents, they did not show a statistical significant association between SEP assessed by EDI in quintile and the fact of being non-respondent.”

9. The results of testing the influence of excluding EDI from the imputation model on the model 3 seems out of place at the end of discussion.

The authors’ response: We agree with the reviewer comment and have moved it to the end of the third paragraph of the discussion section.

10. Results: Please report at least OR and p-value for those claims of significant associations throughout the results section.

The authors’ response: In the text, we chose to not repeat the detailed results (OR and P-value) that are already presented in the tables to facilitate the reading of the manuscript. The text accompanying the tables thus yields a summarize of the results which are detailed in the tables.

11. Lines 197 -199 Results: "The proportion of advanced stage cancer…. and 41%." What was the purpose of reporting these %s? And these %s cannot be found anywhere in the tables. Apparently, the authors ran analyses for the distribution of stage at diagnosis and EDI quintile (at least for model 2).

Please also indicate this in the statistical analyses.

The authors’ response: We clarified this sentence by adding “Among the 1218 respondents,”. These results were presented to yield a potential explanation behind the differences of statistical significance for the SEP-outcome associations using the different measures. It did not correspond to supplemental analysis but just the description of the proportion of advanced stage at diagnosis in each stratum of SEP using IDI or EDI. We did not present this in any table to avoid overloading the result section.

12. It was unclear why the authors only compared the results of model 3 with model 1.a but not the results of model 4 with model 1.b.

The authors’ response: Model 1.a. consists in running complete case analysis of the relationship linking the individual deprivation index (IDI) to the stage at diagnosis. Compared these results to
those from model 3 using the imputed version of IDI highlight the contribution of the 426 non-respondents in the estimates. Indeed, although imputation may have introduced a “regression to the mean” effect in the measure of these women’s SEP, the SEP-stage at diagnosis is strengthened by the inclusion of the non-respondents. The same phenomenon is observed when we compared models 4 to model 1.b. In the manuscript the discussion I developed above is summarise by “However, the comparison of the results from the different analyses indicates that the association between SEP and outcome is strengthened by the inclusion of the 426 women with missing data on IDI” (third paragraph of the discussion section, lines 272-274).

13. The differences of ROC were small, and it is important to test the differences on ROC (model 1a vs. model 1b; model 3 vs. model 4). In addition, the ROCs were low and the authors should discuss this issue.

The authors’ response: We respond to this comment in two points. First, we added the following text in the beginning of the third paragraph of the discussion section: “Therefore, we were not surprised by the relative smallness of the AUC because these models may miss some important intermediate and confounding variables in the relation between SEP and stage at diagnosis. Nevertheless, this is not a limitation regarding our objective of discussing the use of ecological-level SEP measure to deal with missing data when individual measures are available but uncompleted.” Second, we have also added the presentation of the formal tests for the comparison between AUC based on the algorithms proposed by Delong, Delong, and Clarke-Pearson (1988) using the STATA command roccomp. This was added in to manuscript in the last paragraph of the result section.

14. Please reword “This illustrates the direction of the selection bias that occurs when women who did not respond to the self-reported questionnaire are excluded.” It is unclear.

The authors’ responses: We reworded the text as follow: “This illustrates the direction of the selection bias that occurs when women who did not respond to the self-reported questionnaire are excluded. This bias may likely lead to an underestimation of the SEP-stage at diagnosis relationship when these women are excluded from the analysis. The non-respondent women are older, with more advanced stage at diagnosis and tended to be more deprived than the respondent. In this sub-population, the proportion of advanced stage cancer from EDI quintile 1 to 5 were respectively 37%, 52%, 51%, 61% and 66% (p-value<0.01; Chi-square test with 4 df, data not shown).”
15. Lines 220-221. Please indicate data not shown here.

The authors’ response: This was done at the end of the parentheses.

16. Please report the associated p-values for the analysis excluding EDI from the imputation model.

The authors’ response: This was done and modified in the manuscript.

17. Just put AUC=0.xxx in the table, as you have introduced this acronym earlier.

The authors’ response: This was done.

18. Please provide the references for those 7 papers and 2 papers where DI was used to characterize the contextual environment alone or in complement of individual-level measures of SEP, respectively.

The authors’ response: We updated the manuscript regarding the number of papers dealing with DI to characterize the contextual environment to cover the period from 2012 to the first quarter of 2019 (see the second paragraph of the discussion section). The corresponding references are given in supplemental file.

19. Please provide references to support the statement "However, the indices were more efficient….. and are suitable for measuring deprivation but not affluence."

The authors’ response: This statement refers to the reference 17. We repeated the reference to clarify this point (line 287).

20. Line 280 - Please correct the p-value "](p<0001)" reported here and be more specific about what were "these variables are correlated". The spearman rho was the correlation of EDI with what (and IDI?)?

The authors’ response: It referred to individual SEP, i.e. IDI. We have thus clarified this in the text (fourth paragraph of the discussion section (line 296).
21. No connection of the findings of this study with existing research were discussed.

The authors’ response: Our study was shaped to be a discussion around the use of the ecological measures of SEP to deal with uncompleted in individual-level SEP measures. To our knowledge, we found no such discussion in the literature. However, the results we observed were not surprising as they fall in what was theoretically expected and described in the methodological published works regarding missing data. These connections were done at the end of the discussion section (last paragraph): “Finally, our results are in line with previous study that supported the interest of multiple imputation regarding complete case analysis, in particular in the case of informative missingness (18, 27, 28). We also highlight the possibility that may offer ecological SEP measure for assessing selection bias due to incomplete data on individual-level measures.”

22. No limitations were discussed.

The authors’ response: This was corrected at the beginning of the second paragraph of the discussion section. In this paragraph, the section “Since its development in 2012, almost forty papers using the European Deprivation Index (EDI) was published up to the first quarter of 2019. In more than half of these papers, no individual data were available, and the EDI was used as a proxy of individual-level SEP. In the others, EDI was used to characterize the contextual environment alone (12 papers) or in complement of individual-level measures of SEP (4 papers, see supplemental file). To our knowledge, this is the first time the EDI is used to impute for missing individual SEP.” was moved at the end.