PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Association between socioeconomic status and dental caries among Chinese preschool children: a cross-sectional national study
AUTHORS	Zhang, Tingting; Hong, Jialan; Yu, Xueting; Liu, Qiulin; Li, Andi; Wu, Zhijing; Zeng, Xiaojuan

VERSION 1 – REVIEW

REVIEWER	Viana, Karolline Alves Universidade Federal de Goiás
REVIEW RETURNED	26-Oct-2020

GENERAL COMMENTS	It is an interesting and well-done study that investigates the association between socioeconomic status and dental caries among Chinese preschool children. The study possess just some minor points that require clarifications, as follows: 1) Introduction - Please justify your study better. Why is it important? Just mentioning a “lack of representative data on oral health inequalities for Chinese preschool children” and that “it is needed” is insufficient. 2) Methods - Althought full details of the survey’s design can be found in another paper, some information are needed. Please specify how was the training and calibration process (page 5, lines 37-38) - Please clarify how the information about dental pain was obtained. - Why was the education level divided into 3 groups? Why were these grouping chosen? - And about the home income? Why were these grouping chosen? - Please, to make comparison possible, use a conversion rate of $1 US commercial dollar at the time of data collection. - Please clarify how the information about self-perceived general health was obtained. 3) Discussion - Page 9, lines 41-51: please add a critical reflection. Why do you think these studies have found no association? - Page 11, lines 16-17: What are the advantages of using SII and RII? What information do these indexes add?

REVIEWER	Xiangqun Ju The University of Adelaide, Australia
REVIEW RETURNED	27-Oct-2020

GENERAL COMMENTS	General comments The study aimed to assess the association between child caries status and socioeconomic status (SES) in China. The manuscript is useful, and of interest to readers. However, there were some significant problems with statistical analysis. In addition, the English is not quite there, and there are significant general changes needed. Get someone else (a native English speaker) to check and amend it. To enhance the reader's understanding, it would be helpful for the authors to address the following questions: Abstract 1. The aim of the study in the objective differs from the aims as stated at the end of the introduction Introduction 1. Please add full name for 'GDP' (line 45, Page 3) when it was appeared at the first time Methods 1. Data source: 1) Need more details for clinical dental examination: a) Did the data estimate inter-examiner reliability? b) Did the data estimate the intraclass correlation coefficient (ICC) for untreated decay/dmft? If yes, please report them. 2) 'face to face interview'? (line 34, Page 5) 2. Provide more details for covariates (line 16-18, page 5), such as how did they were categorised, make sure they are consistent with the results, especially consistent with Table 1. 1) offer a little more information of the questionnaires, such as, how did you ask children's 'Self-perceived general health'? and what were answers? 2) use 'boy/girl' to instead of 'male/female' throughout in the text when appropriate. 3) Add variable 'ethnicity' and its classification. 3. Statistical analysis: This section needs reorganize and re-write. I recommend that a biostatistician should be consulted to ensure the results can be generated from the methods you introduce here. 1) Add more details on your multiple imputation. Did MI include outcome variables? 2) The sentences were confused: a) 'Model 1 was adjusted for children's age… exclude the effects of all covariates' (line 49-51, page5). What did 'exclude the effects of all covariates' mean? Was Model 1 adjusted for all covariates or not? b) 'Confounding can lead to an overestimate… Therefore, the effects of the confounding variables…' (line 53-60, page 5). What were confounding variables? Did they seem as 'covariates'? c) It was unclear: did you use weighted data or MI data? (line 13-14, Page 7) d) What was RII estimated? Mean or rate? 3) It was confused: how did you calculate incidence rate ratio for 'dmft'? Did you calculate incidence rates for each age group using person-time at risk? If not, it should be 'mean ratio'. In addition, how did you calculate incidence rate ratio for 'dental pain' and 'Untreated caries'? Results

	1. Don't repeat data, only report the total number (line 46, page 6) 2. Where was the result presented: 'nearly half of children (49.83%) resided…'? (line 50-52, page 8). What did 'resided in home' mean? Did some children not reside in their home? 3. They were not stronger relationship (line 21, page 10), as there was only 10% or 16% (or 1.10 or 1.16 times) higher… 4. Table 1: 1) Add a row to report 'total number , percentage, prevalence and mean' in the Table 1 2) Suggest to report 95% CL in Table 1 3) It was unclear on the footnotes: were 'Frequencies' weighted or 'percentage or prevalence, even mean of dmft' weighted? Discussion 1. The possible deeper causes/reasons should be discussed in each theme/paragraph (Second to Second paragraphs, page 8-9) 2. Add more possible interactions between education level and household income (Last paragraph, page 9) 3. Please provide references; 1) 'supporting the previous literature' (line 30-32, page 8) 2) 'The dmft in Chinese preschool children…3.5-3.35 slightly' (line 31-33, page 9). 4. Please use 'Chinese children aged 3-5 years' throughout in the text when appropriate, don't just use word 'children' (line 15, page 9).

REVIEWER	Enrico di Bella University of Genoa, Italy
REVIEW RETURNED	22-Nov-2020

GENERAL COMMENTS	I have just finished reading your work "Association between ...", and I find it well written and well structured, although I think there are some elements that you should clarify through a revision of it. First of all, I make a general reflection on the innovative nature of your work. The link between the socio-economic condition of families and the health of their members is a widely debated issue. Generally, with some exceptions, this link is significant and positive. However, the causal link between these two aspects is often unclear, in the idea that there is a positive correlation between them but not necessarily a direct dependency relation. This justifies the sometimes non-significant results of dependence among health and education conditions or family income, which you discuss in the paragraph "Discussion". This aspect would, however, merit some consideration because you use both the two variables as covariates of your model. That said, it seems to me that your work produces somewhat predictable results resulting in a "confirmation study" of a large part of the existing literature. Its most innovative part is probably its scope of application. I have no particular observations to make for the first part of the paper, which is well written. I would suggest you consider moving to the introduction some notes referring to previous studies mentioned in the discussion. Concerning the description of the data, I find that some parts are not totally clear. In particular, when you define the three "study measures" you confuse measurements on the individual (dental

	pain experience) with metrics built on all statistical units (prevalence). Please make this part homogeneous. Moreover, the covariates are not even briefly described in the text, apart from "place of residence" which is described even before being mentioned in the text (page 5, lines 12-14). I suggest you to briefly describe each covariate in the same way you did for the two SES variables. Concerning the methodological part, the analyses you propose are standard but adequate. However, I have not found in table 1 the data concerning the output variables: dmtf, dental pain and untreated caries. There are average values for the groups of subjects, but no information on their distributions. For example: how many patients have dmtf = 0? This figure is quite important because if they were consistent in number, you could evaluate the use of Zero-inflated models. I would therefore ask you to add this information and to assess whether it would be appropriate to introduce this aspect into your models. I also do not find the explanation of the multiple imputation method very clear, especially when you mention 20 datasets that have been generated. I would ask you to clarify this aspect, which is discussed later in the paper. Also, regarding the models, I would ask you to verify the absence of collinearities between the variables, especially between the parents' level of education and family income. Finally, in the discussion and conclusions, some aspects jumped out at me. In the various examples that you report from previous studies on the Mongolian, Australian, Chinese, Hong Kong, Mexican and Brazilian populations, you find that the causal link between children's oral health and SES variables is very "heterogeneous". Firstly, I think this discussion can be well placed in the first part of the paper. Secondly, I wonder which of the case studies you mention is the one that comes the closest to yours. As a minor remark, I invite you to carefully re-read the paper which has some typos such as "Ethnics Committee" (page 4 line 41) which I assume is "Ethics Committee" and "Incident Rate" (page 6 line 3) which I assume is "Incidence Rate". I also think that the child's self-perceived general health level should be "Good or Better" and "Fair or Less" rather than "Good and Better" and "Fair and Less".

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Comment 1:
It is an interesting and well-done study that investigates the association between socioeconomic status and dental caries among Chinese preschool children. The study possess just some minor points that require clarifications, as follows:
1) Introduction
- Please justify your study better. Why is it important? Just mentioning a “lack of representative data on oral health inequalities for Chinese preschool children” and that “it is needed” is insufficient.

Reply: Thanks for your comments. This has now been added, See in the Introduction section, highlighted in yellow (page 3, line 24-27).

Comment 2:
Methods
- Although full details of the survey's design can be found in another paper, some information are needed. Please specify how was the training and calibration process (page 5, lines 37-38)
- Please clarify how the information about dental pain was obtained.
- Why was the education level divided into 3 groups? Why were these grouping chosen?
- And about the home income? Why were these grouping chosen?
- Please, to make comparison possible, use a conversion rate of $1 US commercial dollar at the time of data collection.
- Please clarify how the information about self-perceived general health was obtained.

Reply: Suggestions followed and revisions done. See in the Methods section, highlighted in yellow (page 4, line 13-17, line 21 and line 24-29).

Comment 3:
Discussion
- Page 9, lines 41-51: please add a critical reflection. Why do you think these studies have found no association?
- Page 11, lines 16-17: What are the advantages of using SII and RII? What information do these indexes add?

Reply: These has now been added. See a critical reflection in the Discussion section, highlighted in yellow (page 8, line 13). And see the advantages in the Methods section, highlighted in yellow (page 5, line 28).

Reviewer 2

Comment 1:
Abstract
1.The aim of the study in the objective differs from the aims as stated at the end of the introduction.

Reply: Revision done. See in the Abstract section, highlight in yellow (page 2, line 4).

Comment 2:
Introduction
1.Please add full name for 'GDP' (line 45, Page 3) when it was appeared at the first time.

Reply: 'GDP' has now been dropped in the revision.

Comment 3:
Methods
1. Data source:
1) Need more details for clinical dental examination:
a) Did the data estimate inter-examiner reliability?
b) Did the data estimate the intraclass correlation coefficient (ICC) for untreated decay/dmft? If yes, please report them.
2) 'face to face interview'? (line 34, Page 5)

Reply: These have now been added except for ‘face-to-face interview’. See in the Methods section, highlighted in yellow (page 4, line 13-17). ‘Untreated caries’ has been dropped in the revision considering for the intraclass correlation coefficient (ICC) of untreated caries and dmft was 0.987.

Comment 4:
Provide more details for covariates (line 16-18, page 5), such as how did they were categorized, make sure they are consistent with the results, especially consistent with Table 1.
1) offer a little more information of the questionnaires, such as, how did you ask children’s ’Self-perceived general health’? and what were answers?
2) use ‘boy/girl’ to instead of ‘male/female’ throughout in the text when appropriate.
3) Add variable ‘ethnicity’ and its classification.

Reply: Details for covariates have now been added in the revision, which can be seen in the Methods section, highlighted in yellow (page 4, line 13-17, and page 5, line 3-5).

Comment 5:
3. Statistical analysis:
This section needs reorganize and re-write. I recommend that a biostatistician should be consulted to ensure the results can be generated from the methods you introduce here.
1) Add more details on your multiple imputation. Did MI include outcome variables?
2) The sentences were confused:
a) ‘Model 1 was adjusted for children’s age… exclude the effects of all covariates’ (line 49-51, page5). What did ‘exclude the effects of all covariates’ mean? Was Model 1 adjusted for all covariates or not?
b) ‘Confounding can lead to an overestimate… Therefore, the effects of the confounding variables…’ (line 53-60, page 5). What were confounding variables? Did they seem as ‘covariates’?
c) It was unclear: did you use weighted data or MI data? (line 13-14, Page 7)
d) What was RII estimated? Mean or rate?
3) It was confused: how did you calculate incidence rate ratio for ‘dmft’? Did you calculate incidence rates for each age group using person-time at risk? If not, it should be ‘mean ratio’. In addition, how did you calculate incidence rate ratio for ‘dental pain’ and ‘Untreated caries’?

Reply: More details of Statistical analysis have now been added in the revision. See in the Statistical analysis section, highlighted in yellow (page 5, line 13-27, and page 6, line 1-10).

Comment 6:
Results
1. Don’t repeat data, only report the total number (line 46, page 6)
2. Where was the result presented: ‘nearly half of children (49.83%) resided…’? (line 50-52, page 8). What did ‘resided in home’ mean? Did some children not reside in their home?
3. They were not stronger relationship (line 21, page 10), as there was only 10% or 16% (or 1.10 or 1.16 times) higher…
4. Table 1:
1) Add a row to report ‘total number, percentage, prevalence and mean’ in the Table 1
2) Suggest to report 95% CL in Table 1
3) It was unclear on the footnotes: were ‘Frequencies’ weighted or ‘percentage or prevalence, even mean of dmft’ weighted?

Reply: Suggestions and revisions done. See in the Results section, highlighted in yellow (page 6, line 18-22). As for Table 1, new revisions can be seen, highlighted in yellow.

Comment 7:
Discussion

1. The possible deeper causes/reasons should be discussed in each theme/paragraph (Second to Second paragraphs, page 8-9)
2. Add more possible interactions between education level and household income (Last paragraph, page 9)
3. Please provide references;
1) ‘supporting the previous literature’ (line 30-32, page 8)
2) ‘The dmft in Chinese preschool children…3.5-3.35 slightly’ (line 31-33, page 9).
4. Please use ‘Chinese children aged 3-5 years’ throughout in the text when appropriate, don’t just use word ‘children’ (line 15, page 9).

Reply: Suggestions and revisions done. See in the Discussion section, highlighted in yellow (page 8-9).

Reviewer 3
Comment 1:
I have just finished reading your work "Association between ...", and I find it well written and well structured, although I think there are some elements that you should clarify through a revision of it.

First of all, I make a general reflection on the innovative nature of your work. The link between the socio-economic condition of families and the health of their members is a widely debated issue. Generally, with some exceptions, this link is significant and positive. However, the causal link between these two aspects is often unclear, in the idea that there is a positive correlation between them but not necessarily a direct dependency relation. This justifies the sometimes non-significant results of dependence among health and education conditions or family income, which you discuss in the paragraph "Discussion". This aspect would, however, merit some consideration because you use both the two variables as covariates of your model.

That said, it seems to me that your work produces somewhat predictable results resulting in a "confirmation study" of a large part of the existing literature. Its most innovative part is probably its scope of application.

I have no particular observations to make for the first part of the paper, which is well written. I would suggest you consider moving to the introduction some notes referring to previous studies mentioned in the discussion.

Reply: Thanks! Suggestions followed and revision done.

Concerning the description of the data, I find that some parts are not totally clear. In particular, when you define the three "study measures" you confuse measurements on the individual (dental pain experience) with metrics built on all statistical units (prevalence). Please make this part homogeneous.

Reply: This has now been added. See in the Methods section, highlighted in yellow (page 4, line 21).

Comment 2:
Moreover, the covariates are not even briefly described in the text, apart from "place of residence" which is described even before being mentioned in the text (page 5, lines 12-14). I suggest you to briefly describe each covariate in the same way you did for the two SES variables.

Reply: This has now been added. See in the Covariates section, highlighted in yellow (page 5, line 3-5).

Comment 3:
Concerning the methodological part, the analyses you propose are standard but adequate. However, I have not found in table 1 the data concerning the output variables: dmft, dental pain and untreated caries. There are average values for the groups of subjects, but no information on their distributions. For example: how many patients have dmft = 0? This figure is quite important because if they were consistent in number, you could evaluate the use of Zero-inflated models. I would therefore ask you to add this information and to assess whether it would be appropriate to introduce this aspect into your models.

Reply: We also considered Zero-inflated negative binomial regression. Since the proportion of “zero” dmft counts was 37.5%, negative binomial regression was used to analysis dmft. The explanation has now been added in the Statistical analysis, highlighted in yellow (page5, line 21). And total number of dependent variables can be found in Table 2.

Comment 4:
I also do not find the explanation of the multiple imputation method very clear, especially when you mention 20 datasets that have been generated. I would ask you to clarify this aspect, which is discussed later in the paper.

Reply: More details of Multiple imputation (MI) method have now been added in the Statistical analysis section, highlighted in yellow (page 5, line 12-16).

Comment 5:
Also, regarding the models, I would ask you to verify the absence of collinearities between the variables, especially between the parents' level of education and family income.

Reply: Suggestions and revisions done. See in the Statistical analysis section, highlighted in yellow (page 5, line 16-19).

Comment 6:
Finally, in the discussion and conclusions, some aspects jumped out at me. In the various examples that you report from previous studies on the Mongolian, Australian, Chinese, Hong Kong, Mexican and Brazilian populations, you find that the causal link between children's oral health and SES variables is very "heterogeneous". Firstly, I think this discussion can be well placed in the first part of the paper. Secondly, I wonder which of the case studies you mention is the one that comes the closest to yours.

Reply: Suggestions and revisions done. See in the Introduction section, highlighted in yellow (page 3, line 10-19). Actually, our study is not very similar to any of the studies mentioned above, because the results show that different parental educational attainment and different household income level show different trends of inequality in different dental caries indicators by places of residence.

Comment 7:
As a minor remark, I invite you to carefully re-read the paper which has some typos such as "Ethnics Committee" (page 4 line 41) which I assume is "Ethics Committee" and "Incident Rate" (page 6 line 3) which I assume is "Incidence Rate". I also think that the child's self-perceived general health level should be "Good or Better" and "Fair or Less" rather than "Good and Better" and "Fair and Less".

Reply: Revisions Done. See in the Methods section, page 4, line 10-11, highlighted in yellow. And child general health was revised. See in the Methods section, page 5, line 4-5, highlighted in yellow.

VERSION 2 – REVIEW

REVIEWER	Viana, Karolline Alves Universidade Federal de Goiás
REVIEW RETURNED	29-Jan-2021

GENERAL COMMENTS	I greatly appreciate all modifications, which I truly believe contributed toward improving the text. The manuscript has been revised according to the reviewer's questions.

REVIEWER	Ju, X University of Adelaide, Adelaide Dental School
REVIEW RETURNED	05-Feb-2021

GENERAL COMMENTS	General comments The authors have addressed my previous comments satisfactorily, and made improvements in the manuscript. Minor point 1. What did the word 'weighted' mean from the sentence 'Taking into account …, all estimates were weighted' (line 18, page 7)? Did you use both MI and weighted data to do statistical analysis? if yes, your results maybe overestimate. OR the 'weighted' means using MI data to do analysis only? 2. It was difficult to read the Figure 1. Which sub-figures were for parental education or Household income? Please add labels in Figure 1. 3. What does '0.20' (the red colour) mean? Is it '-0.20'?

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1

Dr. Karolline Alves Viana, Universidade Federal de Goiás

Comment to the Author:

I greatly appreciate all modifications, which I truly believe contributed toward improving the text.

The manuscript has been revised according to the reviewer's questions.

Reply: Thanks for your comments.

Reviewer 2

Dr. X Ju, University of Adelaide

Comments to the Author:

General comments

The authors have addressed my previous comments satisfactorily, and made improvements in the manuscript.

Reply: Thanks for your comments.

Comment 1

What did the word ‘weighted’ mean from the sentence ‘Taking into account …, all estimates were weighted’ (line 18, page 7)? Did you use both MI and weighted data to do statistical analysis? if yes, your results maybe overestimate. OR the ‘weighted’ means using MI data to do analysis only?

Reply: We did statistical analysis by both MI and weighted data. However， only the results based on weighted data was showed in the paper since the difference of results from unweighted data and weighted data analysis was very small and also MI can increase precision and reduce bias. And the revision was done. See in the Method section, page 6, line16.

Comment 2

It was difficult to read the Figure 1. Which sub-figures were for parental education or Household income? Please add labels in Figure 1.

Reply: Revision done. See in the Figure 1.

Comment3

What does ‘0.20’ (the red colour) mean? Is it ‘-0.20’?

Reply: Revision done. See in the Figure 2.