Reviewer’s report

Title: Vitamin D status and dental caries in healthy Swedish children

Version: 0 Date: 06 Oct 2017

Reviewer: Geraldine Lo Siou

Reviewer's report:

Although not a statistical issue, it would have been helpful for the reviewers and readers to present a flowchart of the study design. It was confusing to understand what was measured at different time points: (1) baseline/intervention - when children were 6 years old, (2) three months after intervention, (3) two years after intervention.

General comments:

* Please present medians and interquartile ranges whenever presenting means and 95% CI for continuous variables in tables.

* If normality assumption is not met, one suggestion is to conduct nonparametric tests (i.e., Wilcoxon rank-sum test, Kruskal-Wallis tests) instead of the Student's t-test and ANOVA.

* A few concerns regarding the use of PCA or PLS: maybe just a misunderstanding of the variables that were used to conduct PCA or PLS. Please provide more details about this analysis.

* Please present results separately by intervention groups in table 1.

Detailed comments:

1. Lines 93-94: It will be helpful to present the proportion of participating children (n=85) who were from Northern vs. Southern Sweden.

2. Line 95: It will be helpful to present the proportion of participating children (n=85) who were fair vs. darker skin types.
3. Lines 98-100: It will be helpful to present all results separately by the three intervention groups as at first visit in DViSUM, children were randomly assigned to receive 2, 10, or 25 μg of vitamin D3 per day in a milk-based supplement for three months.

4. Lines 103-106: "Of the 206 children who participated in DViSUM, only 41% consented to participate in an examination of their dental status […]. The major reason for non-participation was that the children's caretakers had moved out of the catchment areas": It will be helpful to present all the reasons as some reasons may be related to the intervention program.

5. Lines 182-183: Were all continuous variables normally distributed? Very often, dietary intake data are not normally distributed. One suggestion is to present the medians and interquartile ranges as well as the means and 95% CI.

6. Lines 183-184: Normality assumption should be met when conducting Student's t-test or ANOVA. If this assumption is not met, one suggestion is to conduct Wilcoxon rank-sum test or Kruskal-Wallis test instead.

7. Lines 188-190: Data for caries, food intake and vitamin D at baseline and at follow-up (three-months or two-years after intervention?) are repeated measures. Thus, one suggestion is to conduct intra-class correlation coefficients with 95%CI instead of Spearman's correlation coefficients.

8. Lines 192-194: In table 4, was "vitamin D status" used as a stratifying variable instead of an independent variable, as stated in the sentence?

9. Lines 198-200: It seems "number of teeth", "tooth brushing", and "presence or absence of S. mutans" were collected at two-years after intervention; whereas "father's and mother's education levels", and "region of residence" were collected at baseline. If so, please explain how model 1 took into account the fact that those variables were collected at different time points?

10. Line 202: Blood samples were also collected at three-months after intervention, which coincides with February-March 2013. Please confirm that those months are still considered to be in the winter season.

11. Lines 205-206: By "sensitivity analyses by region of residence, skin colour, and vitamin D supplementation at 8y", does this sentence mean stratification analyses?
12. Line 207: minor typo here, did the authors meant "multivariate principal components analysis (PCA)" instead of "multivariate partial least square analysis (PCA)" (as currently written in manuscript)

13. Lines 207-210: There are a few concerns here:

a. Is the objective of using PCA or PLS in this paper for dimensionality reduction? If yes, this objective was not clearly stated. In addition, PLS is used to find linear combinations of the variables to predict response variables linearly … What were the response variables in PLS?

b. Were all variables continuous (i.e., lifestyle measures, medical variables)? What did "lifestyle measures" refer to? Were they variables on diet intake and socio-economic conditions which were collected at baseline? Same questions about "medical variables": did they refer to dental health-related behaviours which were collected at two-years after intervention? If so, there may be a mixture of continuous, categorical and binary variables, was this taken into account when conducting PCA and PLS?

c. Were PCA and PLS regression conducted on variables that were repeated measures (i.e., serum/plasma samples were collected at baseline, at three-months after intervention and at two-years after intervention) and on variables which were collected at different time points (i.e., lifestyle measures collected at baseline; medical variables collected at two-years after intervention)? If so, conducting PCA or PLS on such data is complicated as there are correlations among the same variables collected at different time points, and not all variables were collected at all three time points. In general, PCA or PLS are conducted on independent observations, thus I would question the use of PCA or PLS with repeated measures (i.e., longitudinal data).

14. Lines 226-231: Since children were randomly assigned to three intervention group at baseline, one suggestion is to present the results separately by intervention groups. Especially regarding the mean vitamin levels, as these may be different depending on which intervention group the children belong.

15. Lines 291-292: Was this Spearman's correlation coefficient? If yes, data for serum vitamin D levels at baseline and at follow-up (three-months or two-years after intervention?) are repeated measures. Thus, one suggestion is to conduct intra-class correlation coefficients with 95%CI instead of Spearman's correlation coefficient.

16. Table 1: It will be helpful to present results separately by intervention groups.
17. Table 3: 47 (caries-free) + 36 (caries) = 83, shouldn't this total be 85?

18. Table 4: minor typos here, did the authors mean "Odds Ratios" instead of "β-coefficient"? In addition, did the authors mean "3 months after intervention" instead of "2 months after intervention"?

**Level of interest**
Please indicate how interesting you found the manuscript:

An article whose findings are important to those with closely related research interests

**Quality of written English**
Please indicate the quality of language in the manuscript:

Acceptable

**Declaration of competing interests**
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.
I declare that I have no competing interests

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal