Author's response to reviews

Title: The development of socioeconomic health differences in childhood: results of the Dutch longitudinal PIAMA birth cohort.

Authors:

Annemarie Ruijsbroek (annemarie.ruijsbroek@rivm.nl)
Alet H Wijga (alet.wijga@rivm.nl)
Marjan Kerkhof (m.kerkhof@epi.umcg.nl)
Gerard H Koppelman (g.h.koppelman@bkk.umcg.nl)
Henriette A Smit (H.A.Smit@umcutrecht.nl)
Mariël Droomers (mariel.droomers@rivm.nl)

Version: 2 Date: 17 December 2010

Author's response to reviews: see over
Coverletter Manuscript The development of socioeconomic health differences in childhood: results of the Dutch longitudinal PIAMA birth cohort.

I hereby send you our response to the comments of the reviewers. The comments are addressed point by point. Subsequently we addressed the points made by the editor at the end of this cover letter.

Reviewer: Gwenn Menvielle

1. Important aspects are missing in the method section. The authors state that « determinants are considered explanatory when education health disparities are reduced by 15% or more ». There is no justification for the choice of this cut-off point. More detailed information should be given about the method used to create the imputed dataset. Many readers are not familiar with this method. Which predictors have been used? also, in the text, the authors mention « five imputed datasets » and then state that they « report results from the analyses of the imputed dataset » (dataset is singular, suggesting that they used only one dataset). This should be clarified.

The referent suggests that the method section should include more information about our methodological choices. First, the comment concerning the 15% reduction in OR as cut-off point. Because of other comments made by the second reviewer and the editor, we have decided to change the way we present the results and have omitted the percentages clarification from the manuscript. Therefore, further clarification of the cut-off point is no longer relevant here. See the second comment of the second reviewer for information about the changes made in the manuscript.

Second, the comment concerning multiple imputations, we elaborated more on the technique of multiple imputations of our dataset. In our analyses we used all 5 datasets so we adjusted the sentences in the method section concerning the multiple imputations. We included several references on multiple imputation techniques.

2. The tables should be revised. They should be self-explanatory. Please add footnotes to the tables.

We added footnotes to the tables and confidence intervals in table 4 (table 6 in new manuscript).

3. The discussion part is poor. The authors used maternal education as an indicator of the SES. This choice and its possible implications on the results are not discussed. What if they had chosen a SES measure related to the father, or to the family? When we study health in childhood, the characteristics of the family are relevant, in particular whether the mother is a lone mother. The financial aspect is also relevant. All these issues are not present in the manuscript and should be discussed. Do the authors have information on these aspects in their cohort? If so, did they perform sensitivity analyses? For instance, if the results substantially change when we account for lone mother, then the conclusions also differ in terms of public health recommendations.

We have repeated our analyses with the educational level of the father as the indicator for SES. These results were very similar.

Unfortunately we do not have information on the other factors the reviewer mentions. In case of the household situation (lone mother or not), we only have an indirect question in the first year-questionnaire (number of adults in the household), which can only give an indication. Furthermore, in the first year questionnaire the number of single parents (determined as families with one adult in the household) was very low (2.3%). We have no data on income level either.

We included the following elaboration on our SES indicator in our discussion:

"Maternal educational level is used in this study as indicator for family socio-economic background. Educational level of the mother is a good proxy for social status in The Netherlands. This has also been supported by the literature (van Berkel-van Schaik AB, Tax B, 1990; Oakes & Rossi 2003). It is a predictor for the knowledge people posses and for the income they are likely to gain. Also, we
repeated our analyses with paternal educational level and found similar results. Other indicators for family socio-economic background were not available in this study.”

4. The possible explanation for health differences should be more discussed. They used day-care attendance in the first year. Why did they not consider day-care attendance before the child goes to school?

In our study we focus on the health determinants around birth and in early childhood (first year in life of the child). This is why we also focus on day-care in early childhood. We added this information to the method section. See also our response on the next point (point 5).

5. The explanatory factors are measured at birth or around birth. When we study health until the age of 8, other factors, measured later in the child’s life, could play a role. For instance, the authors do not mention diet or physical activity, which are likely to play a role among children. This may be particularly relevant for overweight and obesity.

With this study we want to examine how biological and lifestyle factors early in life may affect the development of health inequalities. Therefore we have chosen determinants that occur around birth and not at age 5 or 7. All determinants are measured around the first year of the child. We agree that other factors during childhood may influence health disparities also, but that is not our goal here. We clarified this more clearly in the method section where the determinants are described.

6. The measure of smoking is relatively crude. For the mother, it is smoking during the first four weeks of the pregnancy, so basically before the woman knows she is pregnant or until she realizes she is pregnant. In the discussion, they mention the effect of maternal smoking during pregnancy on the child’s health. Do the authors really consider the variable they use as a proxy of maternal smoke during the pregnancy? The other measure of smoking is more than once a week at home. Why did the author chose this definition for environmental tobacco smoke? A more expanded discussion on the tobacco variables is needed.

Smoking during pregnancy is defined as smoking in the first four weeks. After that, some mothers stopped smoking, while others didn’t. This is tested by the researchers of the PIAMA group using data on smoking behavior after child birth. The indicator we used has an effect on health of the child and therefore we conclude that smoking during pregnancy is detrimental for the child. We have added information about the meaning of this indicator to the discussion and adjusted our recommendations accordingly. We suggest that women should stop smoking even before they become pregnant, for instance when they stop with birth control measures.

Furthermore, we defined environmental tobacco smoke as smoking in the house more than once a week versus not. This indicator is used, because it gives information about whether people are smoking regularly in the proximity of the children, namely in the home environment.

Reviewer 2: Henrik Ohlsson

1. In their second aim (possible explanations for these socioeconomic differences in childhood health) the authors use an effect decomposition approach to explore their research question. While this approach may (?) be valid for continuous outcomes, it often requires unjustifiable assumptions (see A further critique of the analytic strategy of adjusting for covariates to identify biologic mediation – Kaufman, J, MacLehose, R, Kaufman, S. Epidemiologic perspectives and innovations Oct 2004). At the least, a discussion of this issue is warranted.

Thank you very much for pointing out this issue. We would be happy to discuss this matter with you here. The decomposition approach is in our opinion a broadly accepted and widely used method to describe the impact of explanatory factors on an association (see for instance: Jansen et al, 2008, explaining educational inequalities in preterm birth, Arch Dis Child Fetal Neonatal Ed; Kamphuis et al., 2008, Socioeconomic status, environmental and individual factors and sports participation, Med
We have once again scrutinized the papers about this written by Kaufman and colleagues. They stress that the assumption of no confounding between the determinant and the outcome as well as the absence of unit-level interaction or synergism need to be satisfied. They state that this is “particularly difficult to assert in a real world analysis”. Likewise, it is particularly difficult to assert that there is confounding or unit-level interaction present. Kaufman et al. do not advice how to check these assumptions. We once more refer to the body of research that applies the same discussed methodology to which we want to add our results to further the discussion.

We are aware of the limitations of the methodology we used, but also Kaufman et al. do not provide any alternatives. We agree that the quantification of the causal indirect effect is questionable and have therefore omitted the percentages clarification from our considerations and manuscript.

2. The use of this effect decomposition approach may be even more problematic when several determinants are included in the model. The authors conclude, for example, that the relation between educational level of the mother and children’s poor general health is almost completely (75%) explained by the biological and lifestyle factors. I have some hesitations regarding this approach – when several mediation hypotheses are each tested with a simple mediator model, these separate models may suffer from the omitted variable problem, which can lead to biased parameter estimates. Moreover, this approach can also lead to the conclusion, if you use enough mediators, that you explain more than 100% of the studied relation. For example on page 9 it is stated that “about one third of the higher prevalence of weight problems among children from lower educated mothers can be related to the biological and lifestyle determinants studied. How about the correlation of the biological and lifestyle determinants?

When numerous mediators are used in a multiple mediation model, the likelihood of parameter bias due to omitted variables is reduced. Furthermore, it is possible to determine to what extent specific variables mediate the relation between maternal education and the outcome, conditional on the presence of other variables in the model. It is important to remember that a specific indirect effect through a mediator (say, breastfeeding) in the multiple mediation context is not the same as the indirect effect through breastfeeding alone, except in the unlikely circumstance that all other mediators are uncorrelated with breastfeeding. Moreover, table 4 includes very little information regarding the different variables included in the model. The authors must address these issues in a revised manuscript.

With this manuscript, we wish to understand which determinant affects and explains the relation between maternal educational level and children’s health. We are aware that by including only one explaining variable in the model, the results not only include the effect of this specific variable on the relation in question, but also everything else that is connected to this variable and the relation between maternal education and the outcome. However, we wish to continue showing these analyses. As mentioned above, we agree that the quantification of the causal indirect effect is questionable and have therefore omitted the percentages clarification from the manuscript. This way, we put less emphasis on the size of the reduction in OR and describe the findings more qualitatively (which determinants are most important for explaining socioeconomic differences in childhood health). To do so, we have changed the structure of the manuscript and now start with a conceptual model (acyclic graph you mention later on). Next, present information on the relations between the determinants and the health outcomes and information on the relations between maternal educational level and the determinants (shown in two new tables). We have excluded those determinants that are
unrelated to either maternal educational level or the health outcomes. This is described in the method section.

Table 4 (table 6 in new manuscript) still shows the effect of the separate determinants on the relation between maternal educational level and health, and the total model includes all determinants relevant to the health outcome in question and maternal educational level. Determinants were excluded from the total model if no reduction of the odds ratios and/or of the 95% confidence intervals of the odds ratios of the different educational groups were found. In footnotes we describe which determinants are included in the total models of each health outcome.

3. Additionally, there is an ongoing discussion in the literature on how to calculate the amount of the mediation effects and uncertainty measures (e.g. use of bootstrapping methods). At the least, a discussion of this issue is necessary. The authors have used 15% as cut-off point to determine if the determinants are explanatory or not. Why did the authors use 15%? Why did not the authors include an uncertainty measure of the mediation effect?

We performed GEE analyses, and this makes it impossible to use the likelihood ratio-test to test the reduction of deviance of educational differences due to the mediator. Therefore, we have reported the reduction in OR to give an indication of the explanatory ability of the variables. We no longer describe the percentage reduction in OR and have removed this information from table 4 (table 6 in new manuscript). Therefore, we have not performed bootstrapping to include an uncertainty measure.

4. Page 8, line 18. Lower educated mothers are less likely to breastfeed their children for more than 16 weeks… I cannot find the results from this analysis in the tables. In the results part there are several results, regarding associations between the biological/lifestyle determinants and education and or the health outcome I cannot find in the tables. It is difficult for the reader to judge the importance of the results if they are not shown.

These data are indeed not shown and has been included in the new version of our manuscript (table 4).

5. Maybe the authors should consider the use of directed acyclic graphs in order to clarify for the reader the different pathways in their data material.

We included this figure in the manuscript.

6. Page 8, line 2 – regarding the phrasing statistically significant: First of all the authors have not defined what they mean with statistically significant. Secondly, as pointed out in several articles the phrase is obsolete and is often misinterpreted. Please rephrase and focus on the implications of the range of values in the 95% confidence interval. This problem is also apparent at page 9, line 10 – However the increased risk for frequent respiratory infections among children with intermediately educated mothers is no longer statistically significant raised. The point estimate changed from 1.17 to 1.12 – the question is not if it is statistically significant but rather if the reduction of the point estimate (and the interval) is relevant or not.

We included footnotes to every table explaining the abbreviations. Furthermore, we have adapted the description of the results by focusing less on statistical significance and describing the results in more qualitative terms.

7. The tables are hastily done. The headings in the table section need much more information. In table 3 a and b are not explained. In table 4 a and * are not explained. No confidence intervals are included in table 4. Moreover, I would like information of the share of the different health outcomes for the different educational levels.

We have adjusted the tables with above suggestions. Further, we have included information of the share of the health outcomes for the different educational levels in table 2. To keep the table orderly, we restricted this information for health at the age of 8 years.
8. The figure legend and figure information are also hastily done. What does OR mean (I know it is odds ratio – however I believe that this information must be included). The error bars round the point estimate - is it a 95 % confidence interval or?

Additional information is included in the figure heading.

9. The abstract does not really represent the manuscript – there are no results presented regarding the possible explanatory determinants.

We adjusted the abstract and included our findings concerning the explanatory determinants.

10. Please include and discuss in the discussion part that almost 35 % of the information regarding obesity and overweight was missing.

We included the following information in the discussion: “35 percent of the data on overweight and obesity was missing. This is considerable. We dealt with possible selection bias due to missing data by multiple imputations of the missing data. This way the risk of selection bias is reduced. Also, this way we could make more efficient use of the data, because all available data was used in the analyses.”

11. I also have considerations regarding the policy implications the authors propose. I think it is too simplistic to state that for example – “our findings emphasize that health professionals should promote breastfeeding, especially among parents with low socioeconomic backgrounds”. First of all the authors have not reported the association between breastfeeding and their health outcomes. Secondly, the authors have definitely not discussed the variation within the group of mothers with low socioeconomic background. Therefore, in order to avoid trivializing a complex process and to increase the likelihood that policy discussions are treated with the seriousness that they deserve I believe that this policy implication might be excluded from the manuscript.

We included the information on the associations between breastfeeding and the health outcomes and between maternal educational level and breastfeeding in two new tables. We think that these results endorse the importance of promoting breastfeeding among low SES mothers and want to keep the recommendation on this topic in the discussion section.

12. Page 5, line 29 – The biological and lifestyle determinants used in the explanatory analyses were selected on theoretical grounds. Please provide references.

The biological and lifestyle determinants that the research group considered most plausible and explanatory were included in the analyses. We included this explanation in the manuscript. In the introduction references are displayed for the determinants.

13. There are some small language mistakes in the manuscript that disturbs the reader – maybe it is a good idea to have the manuscript proofread.

For example:
The second heading on page 8 – Please delete Text for this subsection
Page 8, line 15 – change are almost completely explained… to is almost……
Page 5, line 8, …weight en height – change en to and
Page 5, line 20,……next tot status – change tot to to

We corrected the errors had the manuscript professionally checked.

Remarks by the editor
- was household income available in this survey? A sensitivity analysis (reported in an appendix? and discussed in the paper) would be useful.

Unfortunately we don’t have data on household income and therefore cannot answer to your request. However, educational level of the mother is a good proxy for social status in The Netherlands. This has also been supported by the literature (van Berkel-van Schaik AB, Tax B, 1990; Oakes & Rossi 2003). It is a predictor for the knowledge people posses and for the income they are likely to gain.

- A large part of the article is devoted to the mechanisms. This should be reflected in the abstract.

We have adjusted our abstract.

- Methods: directly indicate that separate health variables were created for all years where data were available.

We included this information in the methods section.

- Clearly, the reviewer is right to say that the approach used for mediation is inappropriate. The percentage reduction of the odds ratio does not reflect the share of the effect mediated when including mediators one by one into the model, because of the correlation between mediators. It is the reason why approaches such as ?path analysis? have been developed. For example, it is not possible here to say that breastfeeding contributes individually to 41% of the relation between education and asthma.

We rewrote the manuscript and focused more on the determinants most important for each health outcome in a qualitative manner, instead of reporting percentage reduction in OR.

- imputation slightly increases the prevalence, indicating that children without health problems are somewhat overrepresented in the original dataset? Is it completely sure that it is the only possible explanation?

We know that the children with more health complaints at the start of the PIAMA study have more missing data in later questionnaires. Therefore, children with poor health are overrepresented in the missings. The sentence has been changed into: “imputation slightly increases the prevalence, indicating that children without health problems are somewhat overrepresented in the complete cases.”

- What was the sample size for the complete case analysis (without missing values)? Perhaps you could report this analysis that you performed as an online appendix?

The number of complete cases differs between the health outcomes. For overweight and obesity the missing data is the highest. If we perform our GEE analyses with all possible predictors for the health outcomes, 65% of the data is used in the analyses. We have chosen to report the analyses on the imputed datasets, because we believe this will give more accurate results. However, we included the prevalence at age 8 before and after imputation for each health outcome in table 1, to give more information about the missing data in our dataset.

- The interaction between time and education was assessed on the multiplicative scale. Was there an interaction when it was assessed on the additive scale (using risk differences rather than odds ratios)? Please clearly justify the use of the multiplicative scale. See Kaufman JS. Interaction reaction. Epidemiology 2009;20: 159-60 for details. Perhaps report the interaction on both scales.

We believe we used the multiplicative interaction appropriately, because it is used specifically to obtain age-specific odds ratios. The multiplicative interaction is commonly used in GEE analyses to extract time-specific odds ratios from the analyses. To prevent confusion, we removed the P-values of the interactions from the results and from table 3.
- You write: “Day-care attendance is associated with poor general health? Is it not the opposite? If it is the case, how to explain that day-care attendance explains the educational differences in poor general health?”

This should be: Day-care attendance is negatively associated with poor general health. We changed the sentence in: “Day-care attendance is negatively associated with poor general health, meaning that children attending day-care facilities have less often poor general health than children not attending day-care facilities, while children from the lowest educated mother less often attend day-care facilities.”

Moreover, we find it very important that you improve the wording and the English in various parts of the article. It will need to be improved.

To improve the English we had the manuscript professionally checked.

Minor aspects:

- Introduction: “An increasing maternal age is related to the educational, behavioural, and mental health??. In which direction?”

Sentence has been changed into “positively related”

- Introduction: briefly explain the sibling effect.

- Please pay attention to the tenses used for the verb: write all of the article, especially the methods section, with past tenses.

ok

- Page 8: “younger mothers more often report asthma symptoms for their children?”. Similarly on page 12: “asthma symptoms and respiratory infections for their children?.

- Page 9: “clarified by the biological and lifestyles?: “clarified” seems inappropriate.

All the above mentioned aspects are addressed in the improved version of the manuscript.

- Please clarify within the Methods section of the manuscript whether you obtained approval to use the data used in this study or whether it is publically available.

We have received approval from the PIAMA research group to use the data. This is clarified also in the methods section.