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)	Maternal caffeine intake during pregnancy is associated with excess growth in infancy and overweight in childhood: results from a large prospective cohort study
AUTHORS	Papadopoulou, Eleni Botton, Jeremie Brantsaeter, Anne-Lise Haugen, Margaretha Alexander, Jan Meltzer, Helle Margrete Bacelis, Jonas Elfvin, Anders Jacobsson, Bo Sengpiel, Verena

VERSION 1 – REVIEW

REVIEWER	DeKun LI Division of Research Kaiser Foundation Research Institute Kaiser Permanente 2000 Broadway Oakland, CA 94612 USA
REVIEW RETURNED	01-Sep-2017

GENERAL COMMENTS	This manuscript addresses an important association between maternal caffeine intake during pregnancy and an increased risk of excessive infant growth and early childhood obesity. The findings from this manuscript provide further support for previous reports with similar findings that pointed to in-utero exposure to caffeine as a potential contributor to childhood obesity risk. The relatively consistent findings from the literature on this topic point to the need to further examine the reported association. The strengths of the study are the large study population and, in addition to an overall association, a relatively consistent dose-response relationship observed for over-growth measured at various ages. While, like any observational studies, there are a few limitations, including those discussed by the authors, those potential limitations, nonetheless, do not overshadow the main findings and conclusion. Thus, the manuscript should be published, which could stimulate more research on this subject.

REVIEWER	Sonja Wehberg Center for Clinical Epidemiology, Odense University Hospital, and Research Unit of Clinical Epidemiology, Institute of Clinical Research, University of Southern Denmark, Odense, Denmark
REVIEW RETURNED	12-Sep-2017

GENERAL COMMENTS	Review comments BMJ-Open-2017-018895 Thank you for the opportunity of reading your article. Obviously, a very large amount of work has gone into this article. It was especially nice to see an application of growth curves, and in general, a bit more advanced statistical methods. There is a lot of material. The paper might benefit from the exclusion of some analyses, since this could facilitate the more detailed presentation of the remaining analyses. Please see my comments in the attached document. I apologize in advance, if some of my comments or questions could have been answered by myself by more extensive reading. Overall comments Comment 1 The paper would benefit from another round of text revision. There are some inconsistencies between parts, especially the sections on outcome measures and statistics need re-writing. Personally, I would have benefitted from a short presentation (a revised Supplementary Figure 3?) of the general analysis strategy, which I believe, is as follows: • Outcome 1 is excess infant weight gain between birth and one year of age based on reported weights (as defined by..). • Outcome 2 is childhood overweight (obesity status) at 3, 5, and 8 years. This outcome is derived in two steps: first, you model the individual growth curves; second, you calculate BMI based on the model-predicted heights and weights and compare to pre-defined cut-offs to determine obesity yes/no. • The effect of caffeine on outcome 1 and 2 were analyzed by (1) logistic regression with covariates XYZ and categorical caffeine intake, (2) logistic regression with covariates XYZ, where caffeine intake was modelled by cubic splines, (3) logistic regression models with specific interaction terms, (4) logistic regression models with different populations (sensitivity analyses), where sensitivity analysis iv) was only performed for outcome 2. • Further outcomes are: weight, height/length, BMI, weight and height gain velocities (from 1 month to 8 years) based on measured or model-predicted measurements (?). These were analyzed by linear mixed models. Comment 2 I would have liked a discussion section with some elaboration on the external validity of your data. With your large sample size, you could comment on the suitability of the WHO birth weight-for-age z-scores in Norway, as well as discuss the reasonability of the proposed cut-offs for obesity at different age-levels. Are 15% of all Norwegian children overweight at 5? And (only) 5% at age 8? This information seems especially relevant for school-age 8, where you have least measurements. Comment 3 Although you correctly state, that observational studies never establish causality, you still would like to do so. I do not want to dismiss your results, but I am still not convinced of causality, and therefore reluctant to accept your (wording of) conclusions. Though you adjusted for many confounders, you could not include all potential ones (for example, actual exercise, gestational diabetes or other medical conditions, parental working status etc.). Gestational weight gain was not selected as a covariate but correlates to caffeine intake (is it really true, that 50% of all mothers gain more than recommended?), and so on. Comment 4 Could you comment on the significance of the results? Assuming causality for the moment, is it correct, that if I reduce my caffeine intake during pregnancy from very high to low, my child at 8 years would weigh 480 gram more on average (Table 3)? And if 100 women reduced their caffeine intake from very high to low, only 12

instead of 18 children are overweight at school ages 6-8 (Supplementary Table 8)? Would you support an intervention approach on these numbers?

Sections

Comments and questions on Study population

- Unfortunately, Supplementary Table 1 (that is, presumably a flowchart on cohort retention) is missing. Especially, I would have liked to see, (1) which gestational ages (weeks) were included, and (2) which covariates were considered relevant for being eligible (since not all covariates in Supplementary Table 4 were included in the regression models, and not all covariates considered in the regression models are presented in Supplementary Table 4).
- Was it possible for mothers to be included more than once in the study population (say, for a child in 2002 and another in 2007)? If so, how were these analyzed?
- The abstract states recruitment from 2002-2009, while here the recruitment period ends 2008?

Comments and questions on Child postnatal growth and overweight

- Some information in this section (for example, estimation details on the growth curves) belongs more to the following section Statistical analysis.
- Implausible measurements: by exclusion, you mean sat to missing, right?
- Details on anthropometric measurements are found in Supplementary Table 3, not 1.
- The sentence "A z-score of >0.67" etc. does not seem quite right. You surely mean z-score difference? And crossing of percentile bands maybe?
- The comments on handling birth weight in the analysis should be gathered in one place (some reference to birth weight also in Statistical analysis.
- Could you comment on the use of correlation to evaluate model fit? Is this the usual approach with growth curves? I would expect high correlation with any half-reasonable model.
- Could you comment on the inclusion of mother-child pairs with at least one postnatal measurement? What does the model-predicted curve look like with one or two measurements only? I would be nice to see some examples (for example, a curve for a child where only one measurement was available, versus a child, where all eleven measurements were reported).
- The sentence Used BMI cut-offs.. should end are presented in Supplementary Table 2, not 3.

Comments and questions on Statistical analysis

- The whole section would benefit from re-structuring.
- Covariate selection: I take it, that by bivariate analysis, you mean fitting a logistic model with outcome 1 (excess infant weight gain) and maternal caffeine intake as well as one other covariate from a list. You then chose to include only covariates whose (overall) p-value was..? The covariate list in the text does not tally with the variables shown in Supplementary Table 4. And variables in Supplementary Table 4 are not all used for modelling. On the other hand, gestational age is adjusted for (see for example footnote to Supplementary Table 6), but is not included in Supplementary Table 4.
- What was the full list of covariates that you checked for inclusion?
- How were the covariates included in the models (as categorized in Supplementary Table 4)?
- The sentence on covariate selection states only excess growth – are the covariates for outcome 2 (childhood overweight) chosen in a similar fashion?
- The extra model for outcome 2, where you adjust for birth weight (Supplementary Table 5), is not described here.
- The extra sensitivity analysis for outcomes 1 and 2, where

	you investigate different sources of caffeine (Supplementary Table 6), is not described here. • Further outcomes / linear mixed models: It is not clear, if these were based on the measured or predicted values? I presume that you used the repeated (actual) measurements as outcome, a random intercept (for the children), random or fixed effect for age (?) and fixed effects for all other covariates? What do you mean by random intercept and slope for weight, ...? • Out of curiosity: why not simply use linear mixed models as growth curves? • It is not immediately clear, for which outcomes the interaction analyses and sensitivity analyses were performed – I guess, outcomes 1 and 2? • You state the following numbers: complete case analysis on N=30,338 for outcome 1 and N=50,943 with complete covariates and at least one postnatal measurement for outcome 2. Which covariates had to be non-missing – all, that you tested for selection, or all, that you included in the models? For outcome 1, there had to be birth weight and reported weight at one-year? Or did you impute one-year weight from the growth curves when missing? For outcome 2, was birth weight required to be non-missing (since only included in part of the models)? • Could you comment on the choice of cubic splines: a piecewise linear effect of caffeine intake might have been easier to communicate? With cubic splines: are the two tail terms linear? Comments and questions on Results • I appreciate your use of adjusted odds instead of, say, higher risk, in describing the results. • Further outcomes / growth up to 8 years: I would have liked to see results on the random term(s). Comments and questions on Tables and Figures as well as Supplementary material • Figure 1: Supposedly, the same children are included at all ages. Could you comment on the apparent drop in overweight/obesity prevalence from 3 and 5 to 8 years? Were most of these children "cured"? More to the discussion: are there data on the distribution of childhood obesity in Norway? Are the 15% at 5 years (in the study) comparable to findings in the general population? • Figure 2: Are point-wise confidence intervals presented? Would prediction intervals be meaningful? • Table 3: I would prefer to state Mean weight instead of Weight etc. • Supplementary Table 2: Are the presented prevalence numbers based on the study data, or were they published together with the cut-offs? In the former case, would you comment on the suitability of the cut-offs in your population? • Supplementary Table 3: Is the missing structure purely sequential, that is, a mother could only stop to report at one point in time? Was it not possible to have a missing value in between two ages? Other general comments • Outcome 1 (Excess infant weight gain) can occur for all birth weight categories. In your study, is it mostly the small children (or if included, the pre-term babies), that have some catch-up growth to do? Since caffeine intake is possibly linked to lower birth weight, the excess growth could simply be a natural repair? • Relation outcome 1 and outcome 2: are the study children actually following standard weight curves? Does an excess-growth child become an overweight-at-5 child – and what happens between 5 and 8? What is the predictive ability for excess growth on obesity (at an individual level)?

	• Pre-term babies (if included): these would not necessarily be SGA, right? How did you treat their age (actual age or allowing for extra period)?

REVIEWER	Darren Greenwood University of Leeds, UK.
REVIEW RETURNED	16-Oct-2017

GENERAL COMMENTS	bmjopen-2017-018895 Comments for the authors: General comments: 1. MoBa is a superb resource from a highly esteemed research team. 2. The Generation R study demonstrated greater maternal caffeine intake to be associated with higher body mass index from 6 months onwards, slightly dampening the authors' claims to be the first to investigate this in infancy rather than childhood. However, the work is substantial, certainly still novel enough for publication, of very high standard and from a much larger cohort than Generation R. Major points: 3. The statistical methods appear very sound. The growth models are good. The consideration of confounding is good, including what not to adjust for postnatally, though important residual confounding may remain, as the authors acknowledge. The consideration of incomplete data is good. 4. The odds ratios are small, but obesity is common. It would be good to quantify this more clearly when interpreting the results. We always need to say "relative to what?" when we quote relative risks, etc. from cohort studies. 5. Maternal smoking is an important confounder adjusted for in the modelling. However, it is an unusually important confounder, very strongly associated with maternal caffeine intake, and strongly associated with accelerated catch-up infant growth. It is therefore important that this is measured very precisely. Self-reported measures may not be precise enough. Categorisation into ever/ never is even worse and loses any dose-response with the confounder. With odds ratios of the magnitude reported in the manuscript, it is well within the range for some of this to be residual confounding by smoking. Smoking may also keep maternal pre-pregnancy BMI artificially low, so not fully adjusting for maternal BMI. Sensitivity analysis amongst the self-reported never smokers would help reduce some of these doubts (for supplemental material). 6. My main concern is about whether it's just that small babies experience catch-up growth, and this really hasn't got anything more to do with maternal caffeine intake other than it's a predictor of having a small baby. And if so, what does that really add, because we already know that? It appears to be an association without much of an established mechanism related to caffeine as such. Is it caffeine specifically that drives accelerated growth, or is it just that maternal caffeine is associated with smaller babies and all babies born small are more likely to experience accelerated growth in infancy, regardless of maternal caffeine intake? Can we look at this in the caffeine non-consumers by way of sensitivity analysis (possibly for supplemental material)? 7. What about paternal caffeine intake during the pregnancy – is this available? This should have no association with birthweight, conditioning on the confounders. So adjusting for it in the model should not change the estimate for maternal caffeine consumption,

	providing all potential confounders for maternal caffeine have been included. Is there some form of sensitivity analysis that could be conducted using this? It's a bit tenuous, I know, and on reflection the authors may legitimately argue that this is a poor suggestion! But if there were residual confounding from smoking or unmeasurable socio-economic factors, and these were partially common to both partners, this might tease some of that out (possibly for supplemental material). Minor points: 8. Reference 4 could be expanded to cite some of those other authorities from a range of countries the authors refer to. 9. Page 5, line 40. Presumably the mother completes the food frequency questionnaire, not the infant. 10. Page 6, line 21 onwards. Presumably some of these units are mg/day.

VERSION 1 – AUTHOR RESPONSE

Manuscript ID bmjopen-2017-018895

Response to the editor's and reviewers' comments

Comments from the Associate Editor:

#1. I agree with Wehberg that the authors need to tone down their conclusions (including in the abstract) and I am not sure they get anywhere by saying the study fulfils some of the causality criteria. Nutritional epidemiology studies have to be interpreted very cautiously in light of the self-reported exposure and residual confounding (Greenwood and Wehberg are very clear about this). The authors need to adopt a far more cautious tone.

Our response: According to the editor's and reviewers' comments we have now revised the sentences of the conclusion in the abstract and in the main manuscript (page 2, line 57 & lines 59-60 & page 16, lines 451-452, as follows:

Conclusion in abstract: "Any caffeine consumption during pregnancy is associated with excess infant growth and increased risk of overweight, mainly at pre-school ages. Maternal caffeine intake can may modify overall weight growth trajectory from birth to 8 years. This study adds supporting evidence for the current advice to reduce caffeine intake during pregnancy and indicate that complete avoidance might be advisable."

Main conclusion: "Our findings not only support the recommendation to limit caffeine intake during pregnancy (<200mg/day) but also indicate that complete avoidance might be advisable."

#2. They might also want to discuss what the paper adds enough to this systematic review:

Food Chem Toxicol. 2017 Apr 21. pii: S0278-6915(17)30170-9. doi:

10.1016/j.fct.2017.04.002. [Epub ahead of print] Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children. https://www.ncbi.nlm.nih.gov/pubmed/28438661

Our response: In this systematic review, the authors have assessed the body of evidence on maternal caffeine intake during pregnancy at a comparator of 300mg/day and the effects on several outcomes including fetal growth and spontaneous abortion. Postnatally, child carcinogenicity and behavior were

also included. However, the outcome included in our study, growth and weight status, were not included. After critically assessing the body of evidence, the conclusion was that a consumption of up to 300 mg caffeine/day in healthy pregnant women is generally not associated with adverse reproductive and developmental effects.

Our study findings are adding to the evidence of effects of maternal caffeine intake during pregnancy on the child's postnatal growth and development of overweight. We have also controlled for the "pregnancy signal", through adjustment for nausea and vomiting, that is, according to the authors of the review, important to do when assessing the causal relationship between maternal caffeine intake and health outcomes. Our findings are in agreement with the previous studies assessing a similar hypothesis, with associations reported in caffeine levels below and above the comparator, indicating that this 300mg/day might not be a safe level when growth is under study. Hence, more evidence is needed for the association between prenatal caffeine exposure and postnatal growth and an updated future critical assessment of these studies. This information is summarized and is included in the discussion of the revised manuscript (page 15, lines 416-424).

#3. Here's another paper they may want to cite:

Pediatr Res. 2017 Jul;82(1):19-28. doi: 10.1038/pr.2017.70. Epub 2017 May 24. Interaction between maternal caffeine intake during pregnancy and CYP1A2 C164A polymorphism affects infant birth size in the Hokkaido study. https://www.ncbi.nlm.nih.gov/pubmed/28355205

Our response: This is an interesting study. We had already discussed the effect modification due to genetic polymorphisms that can modify the ability to metabolize caffeine during pregnancy. Now this study has been added in the discussion section (page 16, lines 439-441), with the following text:

"On the other hand, during pregnancy, maternal caffeine clearance modified the association between maternal caffeine intake and fetal growth restriction, with faster clearance being more detrimental 1. More specifically, a genotype of rapid caffeine metabolism was associated with reduced birth weight while in women with a different polymorphism on the gene CYP1A2 C164A no effect was found 2."

Reviewers' Comments to Author:

Reviewer: 1. Reviewer Name: DeKun LI

Institution and Country: Division of Research, Kaiser Foundation Research Institute, Kaiser Permanente, 2000 Broadway Oakland, CA 94612, USA Please state any competing interests: None

I was the lead author for one of the papers on the same subject published and cited by the current authors.

Please leave your comments for the authors below

This manuscript addresses an important association between maternal caffeine intake during pregnancy and an increased risk of excessive infant growth and early childhood obesity. The findings from this manuscript provide further support for previous reports with similar findings that pointed to in-utero exposure to caffeine as a potential contributor to childhood obesity risk. The relatively consistent findings from the literature on this topic point to the need to further examine the reported association.

The strengths of the study are the large study population and, in addition to an overall association, a relatively consistent dose-response relationship observed for over-growth measured at various ages. While, like any observational studies, there are a few limitations, including those discussed by the authors, those potential limitations, nonetheless, do not overshadow the main findings and conclusion. Thus, the manuscript should be published, which could stimulate more research on this subject.

Our response: We would like to thank the reviewer for his feedback on our manuscript.

Reviewer: 2. Reviewer Name: Sonja Wehberg

Institution and Country: Center for Clinical Epidemiology, Odense University Hospital, and Research Unit of Clinical Epidemiology, Institute of Clinical Research, University of Southern Denmark, Odense, Denmark. Please state any competing interests: None declared

Please leave your comments for the authors below

Thank you for the opportunity of reading your article. Obviously, a very large amount of work has gone into this article. It was especially nice to see an application of growth curves, and in general, a bit more advanced statistical methods.

There is a lot of material. The paper might benefit from the exclusion of some analyses, since this could facilitate the more detailed presentation of the remaining analyses.

Our response: We have taken the reviewers suggestions into careful consideration and we have revised the included material. Please see the detailed revisions as described in the following responses.

Please see my comments in the attached document. I apologize in advance, if some of my comments or questions could have been answered by myself by more extensive reading.

Overall comments

Comment 1

The paper would benefit from another round of text revision. There are some inconsistencies between parts, especially the sections on outcome measures and statistics need re-writing. Personally, I would have benefitted from a short presentation (a revised Supplementary Figure 3?) of the general analysis strategy, which I believe, is as follows:

• Outcome 1 is excess infant weight gain between birth and one year of age based on reported weights (as defined by..). • Outcome 2 is childhood overweight (obesity status) at 3, 5, and 8 years. This outcome is derived in two steps: first, you model the individual growth curves; second, you calculate BMI based on the modelpredicted heights and weights and compare to pre-defined cut-offs to determine obesity yes/no. • The effect of caffeine on outcome 1 and 2 were analyzed by (1) logistic regression with covariates XYZ and categorical caffeine intake,

(2) logistic regression with covariates XYZ, where caffeine intake was modelled by cubic splines, (3) logistic regression models with specific interaction terms, (4) logistic regression models with different populations (sensitivity analyses), where sensitivity analysis iv) was only performed for outcome 2. • Further outcomes are: weight, height/length, BMI, weight and height gain velocities (from 1 month to 8 years) based on measured or model-predicted measurements (?). These were analyzed by linear mixed models.

Our response: We would like to thank the reviewer for this suggestion. We have now revised the statistical analysis section according to the suggested text and have shortened it. Regarding the outcome section, we have tried to revise and shorten it as well (for outcomes: pages 6-7, lines 146-182 and for statistical analysis pages 8-9, lines 189-237).

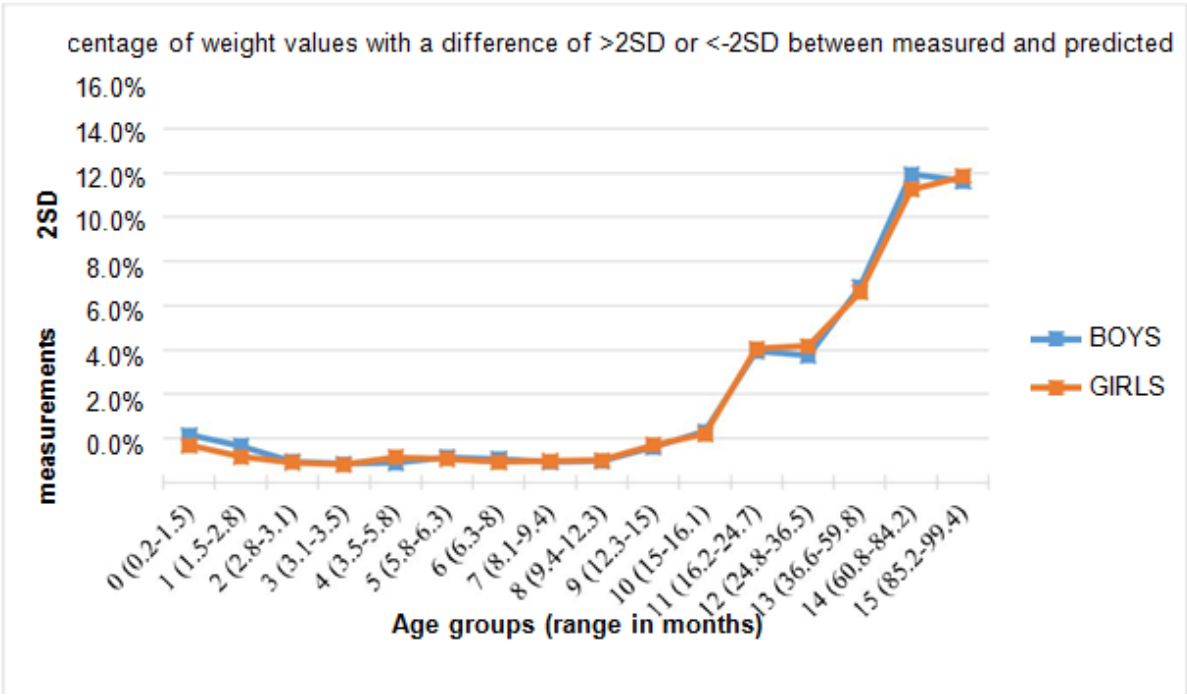
Comment 2

I would have liked a discussion section with some elaboration on the external validity of your data. With your large sample size, you could comment on the suitability of the WHO birth weight-for-age z-scores in Norway, as well as discuss the reasonability of the proposed cut-offs for obesity at different age-levels. Are 15% of all Norwegian children overweight at 5?

And (only) 5% at age 8? This information seems especially relevant for school-age 8, where you have least measurements.

Our response: We clarify that we have used the WHO growth curves as a reference in order to define weight-for-age z-scores and further identify children with excess infant growth. Our choice was based mainly on comparability with other studies. There are several measures of excess growth used in literature, but a gain >0.67 of the WHO-WAZ is the most commonly used measure. We reported a 23% prevalence of excess infant growth from birth to 1 year. Similar prevalence of this measure has been reported before (i.e Ong KK et al : 31%, Karaolis-Danckert N et al: 29%, Valvi et al. :26%, Stratakis et al: mean of 15 European mother-child studies: 35%, range of = 17-56%). Moreover, we have used the IOTF cut-offs to define overweight. This method is based on pre-specific BMI cut-offs.

We cannot explain the "drop" in the prevalence of overweight from 5 to 8 years (from 5 to 8 years: 16% to 4%). This trend was not observed in measured anthropometrics (from pre-school to school age: 14% to 12%) but only in the model-predicted anthropometrics. The percentage of overweight pre-school children who continued being overweight in school age was 40% when using measured anthropometrics and 32% when using predicted anthropometrics (from 5 to 8 years). Therefore, this provides some evidence of agreement between measured and predicted anthropometric data at 8 years. It is possible that this "drop" is a model artifact as this model is modeling growth up until before the adiposity rebound (around 8 years) and weight might have been underestimated in the "end tail" of the growth model. We have calculated the percentage of weight values with a difference of >2SD or <-2SD between measured and predicted for boys and girls (see figure below). This graph shows that approximately after the age of 1 ½ years there are more values with large difference between measured and predicted, compared to younger ages. So poor model fit might explain the low prevalence of overweight as calculated using the predicted measurements. We remind that weight and height measurements with a difference lower than -5SD and higher than 5SD between measured and predicted have been excluded in our analysis and this is described in the methods section.



Comment 3

Although you correctly state, that observational studies never establish causality, you still would like to do so. I do not want to dismiss your results, but I am still not convinced of causality, and therefore reluctant to accept your (wording of) conclusions. Though you adjusted for many confounders, you could not include all potential ones (for example, actual exercise, gestational diabetes or other medical conditions, parental working status etc.). Gestational weight gain was not selected as a covariate but correlates to caffeine intake (is it really true, that 50% of all mothers gain more than recommended?), and so on.

Our response: We agree with the reviewer's comment of not being able to establish a causal relationship through an observational study like ours. We have discussed the issue of unmeasured and residual confounding in the discussion section. Even though the reviewer is not convinced of causality, we have described in the discussion that our results fulfill some of the Bradford-Hill's criteria for causation (i.e a strong association, consistent findings for major caffeine sources, a biological gradient with higher caffeine exposure being associated to abnormal growth, consistent findings in animal models and a plausible mechanism, i.e. fetal programing). In addition, and after following a suggestion from Reviewer #3 we have now conducted negative control analysis using the paternal caffeine intake as a negative control and our findings indicate a minor effect of unmeasured shared familial characteristics (see comment # 7 from Reviewer #3). Finally, the last sentence of our conclusions has been revised now, in a way of not implying a causal relationship between maternal caffeine intake and postnatal growth (see comment #1 from Editor).

Comment 4

Could you comment on the significance of the results? Assuming causality for the moment, is it correct, that if I reduce my caffeine intake during pregnancy from very high to low, my child at 8 years would weigh 480 gram more on average (Table 3)? And if 100 women reduced their caffeine intake from very high to low, only 12 instead of 18 children are overweight at school ages 6-8 (Supplementary Table 8)? Would you support an intervention approach on these numbers?

Our response: Indeed, the observed effects, although significant, can be considered as too low to suggest an intervention. First, as stated above, we cannot conclude about causality. Second, if the effect was causal, as the number of exposed women is high, an intervention could prevent quite a lot of children being overweight, even if the effect is small. Reviewer #3 has made a similar comment; please see our response to #4 from Reviewer #3.

Sections Comments and questions on Study population • Unfortunately, Supplementary Table 1 (that is, presumably a flowchart on cohort retention) is missing. Especially, I would have liked to see, (1) which gestational ages (weeks) were included, and (2) which covariates were considered relevant for being eligible (since not all covariates in Supplementary Table 4 were included in the regression models, and not all covariates considered in the regression models are presented in Supplementary Table 4).

Our response: We have not included a flow chart of cohort retention but a table presenting the available anthropometric measurements by age, and this is supplementary Table 2. We apologize for the mistake and the confusion it might have caused. Regarding the gestational ages of our included population, the range was 28 to 43 weeks of gestation and 90% of the children were born between gestational weeks 37-42. Regarding the tested confounders, maternal height, paternal weight, paternal alcohol consumption and gestational diabetes were considered relevant but did not meet the criteria of inclusion in the models. This information has now been added in the revised manuscript (page 8, lines 209-211). Supplementary Table 4 has also been revised and now includes only the variables that are included in the models as covariates.

• Was it possible for mothers to be included more than once in the study population (say, for a child in 2002 and another in 2007)? If so, how were these analyzed?

Our response: All regression models were adjusted for random effects of sibling clusters since some mothers participated with more than one pregnancy. This information has now been added in the revised manuscript (page 8, lines 200-202).

• The abstract states recruitment from 2002-2009, while here the recruitment period ends 2008?

Our response: The recruitment was finished in 2008, which is correct. Some deliveries occurred in 2009. We apologize for the mistake. This is now corrected in the abstract (page 2, line 40).

Comments and questions on Child postnatal growth and overweight • Some information in this section (for example, estimation details on the growth curves) belongs more to the following section Statistical analysis.

Our response: we would like to keep the description of the methodology on how to apply growth models in the outcome section. This is the first time we have applied such a methodology in the MoBa data and we have considered the predicted growth measurements as a new outcome. Thus, we would like to present it in details in the outcome section.

• Implausible measurements: by exclusion, you mean sat to missing, right?

Our response: Yes this is correct, implausible anthropometric measurements were set to missing.

• Details on anthropometric measurements are found in Supplementary Table 3, not 1.

Our response: We apologize for this mistake. The number of the Supplementary Table is now corrected (page 7, line 161).

• The sentence "A z-score of >0.67" etc. does not seem quite right. You surely mean z-score difference? And crossing of percentile bands maybe?

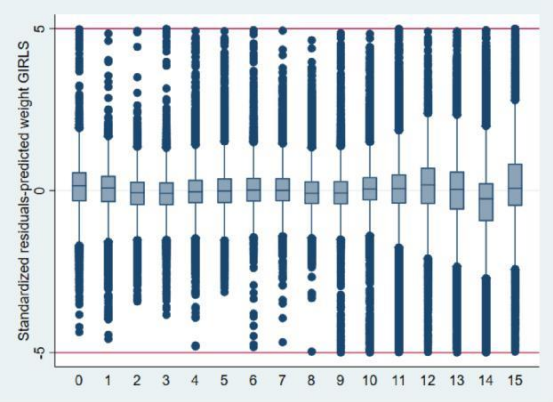
Our response: Yes, this is correct, a gain z-score difference between birth and 1 year>0.67 was used to define excess growth. This is now corrected (page 7, line 165). And we have now added the word “line” next to percentile (page 7, line 165).

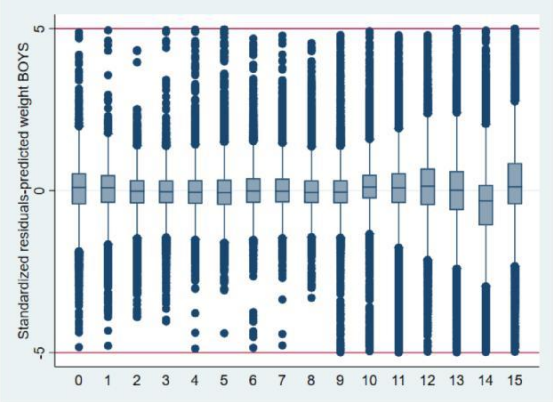
• The comments on handling birth weight in the analysis should be gathered in one place (some reference to birth weight also in Statistical analysis.

Our response: This comment on birth weight is related to the input data for the growth models. We think it is more consistent to present the methodology of growth models in the same section and we would like to keep these sentences as is.

Could you comment on the use of correlation to evaluate model fit? Is this the usual approach with growth curves? I would expect high correlation with any half-reasonable model. • Could you comment on the inclusion of mother-child pairs with at least one postnatal measurement? What does the model-predicted curve look like with one or two measurements only? I would be nice to see some examples (for example, a curve for a child where only one measurement was available, versus a child, where all eleven measurements were reported).

Our response: From a model with good fit we would expect independence between residuals of measured and predicted values and age, for both weight and height. We have provided below the distribution of these residuals, in SD scale, for weight by age 16 age groups in girls and boys. The red horizontal lines at 5SD below and above 0 indicate the cut-offs used for exclusion of implausible measured data. This is described in the methods section. The distribution around 0 is wider for the age groups 12,13, 14, 15 that include measurements at and after the age of 2 years.





We have examined this issue in more details in a previous comment (comment #2 from the same Reviewer) where we calculated that the prevalence of weight values with a difference larger than 2SD or smaller than -2SD is increasing approximately for the ages >1 ½ -2 years. This might indicate a poor model fit after these ages or poor quality of reported anthropometric data.

The application of the growth model with specific examples for different scenarios is very interesting but it is not in the scopes of this study and cannot be added to the manuscript. What we can say is that if a child has 1 measure, the derived trajectory by the model is close to the mean/average growth curve of the population fitted in a way to go through the only point of the child. Indeed this is less precise but it is the best that can be done when a small number of measurements are available. If the children with fewer measurements were excluded from our study, this restriction might have induced selection bias in the estimation of the average growth curve of our population. Therefore, we have included all children in our analysis.

• The sentence Used BMI cut-offs.. should end are presented in Supplementary Table 2, not 3.

Our response: We have revised the numbers of Supplementary Tables in the text.

Comments and questions on Statistical analysis • The whole section would benefit from re-structuring.

Our response: we have now revised the whole section.

• Covariate selection: I take it, that by bivariate analysis, you mean fitting a logistic model with outcome 1 (excess infant weight gain) and maternal caffeine intake as well as one other covariate from a list. You then chose to include only covariates whose (overall) p-value was..? The covariate list in the text does not tally with the variables shown in Supplementary Table

4. And variables in Supplementary Table 4 are not all used for modelling. On the other hand, gestational age is adjusted for (see for example footnote to Supplementary Table 6), but is not included in Supplementary Table 4. • What was the full list of covariates that you checked for inclusion? • How were the covariates included in the models (as categorized in Supplementary Table 4)? • The sentence on covariate selection states only excess growth – are the covariates for outcome 2 (childhood overweight) chosen in a similar fashion?

Our response: The selected variables for adjustment were variables that were associated ($p<0.05$) with both the categorical caffeine intake (4 categories) and excess infant growth (no/yes) as tested by one-way ANOVA, Kruskal-Wallis and chi-square tests. These were: maternal age, maternal and paternal education, parity, pre-pregnancy BMI, paternal BMI, maternal and paternal smoking during pregnancy, maternal energy intake and nausea/vomiting during pregnancy. Gestational age and child's gender were also included in the models as a-priori covariates, and have now been added in Supplementary Table 4. Maternal height, paternal weight, parental alcohol consumption and

gestational diabetes (yes/no) were also considered but not included in the final models as they did not meet the criteria (not presented in Supplementary Table 4).

We decided not to include gestational weight gain in the list of potential confounders as it could be in the causal pathway between caffeine intake and postnatal growth. It was included in the table by mistake. Now Supplementary Table 4 includes only the variables included in the models as covariates. In addition, a detailed description of the variables associated with caffeine intake for the MoBa study has been reported in the previous publication [3] by our research group. Therefore, we did not want to replicate previous findings.

• The extra model for outcome 2, where you adjust for birth weight (Supplementary Table 5), is not described here. • The extra sensitivity analysis for outcomes 1 and 2, where you investigate different sources of caffeine (Supplementary Table 6), is not described here.

Our response: we have now described in detail the sensitivity analyses as well as the analysis of the interaction with birth weight and SGA (page 9, lines 227-237).

• Further outcomes / linear mixed models: It is not clear, if these were based on the measured or predicted values? I presume that you used the repeated (actual) measurements as outcome, a random intercept (for the children), random or fixed effect for age (?) and fixed effects for all other covariates? What do you mean by random intercept and slope for weight, ...? • Out of curiosity: why not simply use linear mixed models as growth curves?

Our response: The outcomes in the linear mixed models were the predicted growth parameters, as derived from the growth model. We have now clarified it in the text (page 8, line 199). As the reviewer has correctly understood, these models take into account that there is a random effect by child (random intercept) and the random effect of age by child is represented as a random slope for age. The effects of the covariates are modeled as fixed. This has now been clarified (page 8, lines 196-200).

Regarding the choice of doing first a structural non-linear mixed-effect growth model rather than a linear mixed model as growth curves, our decision was based on the ability of such models to fit growth data more accurately and to calculate more growth parameters (growth velocity at a given age) with the derivatives of the model.

• It is not immediately clear, for which outcomes the interaction analyses and sensitivity analyses were performed – I guess, outcomes 1 and 2?

Our response: we have now specified the cases when not all the outcomes were assessed. If not specified, then all the outcomes were assessed (page 9, lines 228-238).

• You state the following numbers: complete case analysis on N=30,338 for outcome 1 and

N=50,943 with complete covariates and at least one postnatal measurement for outcome 2. Which covariates had to be non-missing – all, that you tested for selection, or all, that you included in the models? For outcome 1, there had to be birth weight and reported weight at one-year? Or did you impute one-year weight from the growth curves when missing? For outcome 2, was birth weight required to be non-missing (since only included in part of the models)?

Our response: the inclusion criteria are described in the study population section. Mothers and children with available information on SGA status, at least one postnatal measurement, all selected covariates (including birth weight and gestational age) and caffeine intake were the final study population (n=50,943). Of them, 38,338 mother-child pairs were included in the excess infant growth analysis (using WHO-growth reference curves), because they had available anthropometric information at 12 months (age range 9-13 months).

• Could you comment on the choice of cubic splines: a piecewise linear effect of caffeine intake might have been easier to communicate? With cubic splines: are the two tail terms linear?

Our response: We did not pursue to provide piecewise linear coefficients because we did not think this would add to the quality of our evidence. For exposure-levels in which the majority of our population is exposed to, we observed a linear dose-response association by using pre-specified caffeine intake categories (4 categories) and by the graphical representation of the association using cubic splines. And these results were in agreement. In addition, in the cubic splines analysis we tested the linearity of the association by comparing the first term (that models the linear association between the exposure and the outcome) with each of the other terms (non-linear terms) and there was non-significant difference, meaning that the relationship is linear. In the additional sensitivity analysis in which high caffeine drinkers were excluded, the positive dose-response association persisted. Given a linear relationship there is no need to model piecewise effects. Hence, we would like to keep our initial analysis with specific beta presented by pre-defined groups of caffeine intake and with the restricted cubic splines applied to examine the shape of the association.

Comments and questions on Results • I appreciate your use of adjusted odds instead of, say, higher risk, in describing the results. • Further outcomes / growth up to 8 years: I would have liked to see results on the random term(s).

Our response: we have not extracted the variance or the random intercept and the random slope of the models and we have not reported it in the manuscript.

Comments and questions on Tables and Figures as well as Supplementary material • Figure 1: Supposedly, the same children are included at all ages. Could you comment on the apparent drop in overweight/obesity prevalence from 3 and 5 to 8 years? Were most of these children "cured"? More to the discussion: are there data on the distribution of childhood obesity in Norway? Are the 15% at 5 years (in the study) comparable to findings in the general population?

Our response: We have answered this in an earlier comment (comment 2). We have also added the information in the revised manuscript.

• Figure 2: Are point-wise confidence intervals presented? Would prediction intervals be meaningful?

Our response: In figure 2 we have presented the association between caffeine intake and change in weight in our population study. These are estimations based on our data and not predictions. As we understand prediction intervals are the intervals in which a hypothesized child should belong (with a probability of 95%) if the maternal caffeine intake was known. However, this is out of the scopes of our study.

• Table 3: I would prefer to state Mean weight instead of Weight etc.

Our response: The clarification that these effect estimates are the adjusted mean differences of weight, weight velocity and BMI has now been added as a footnote under Table 3.

• Supplementary Table 2: Are the presented prevalence numbers based on the study data, or were they published together with the cut-offs? In the former case, would you comment on the suitability of the cut-offs in your population?

Our response: The prevalence of overweight, including obesity, is based on the BMI defined using predicted anthropometric data in our population.

• Supplementary Table 3: Is the missing structure purely sequential, that is, a mother could only stop to report at one point in time? Was it not possible to have a missing value in between two ages?

Our response: We have not assessed this issue.

Other general comments

• Outcome 1 (Excess infant weight gain) can occur for all birth weight categories. In your study, is it mostly the small children (or if included, the pre-term babies), that have some catch-up growth to do? Since caffeine intake is possibly linked to lower birth weight, the excess growth could simply be a natural repair?

Our response: In our study, we have described this complicated association between prenatal caffeine exposure, restricted fetal growth and accelerated infant weight gain and higher overall weight gain in childhood. It has been shown that neonates with restricted intrauterine growth (i.e SGA, LBW, etc) are more likely to be identified as rapid growers in early life, hence fetal growth is an important mediator in our analysis. However, after excluding SGA neonates and after adjusting for birth weight we obtained similar effect estimates. Hence, our observed associations are not driven solely by the negative effect on fetal growth. Please also see our response to comment #6 by Reviewer #3.

• Relation outcome 1 and outcome 2: are the study children actually following standard weight curves? Does an excess-growth child become an overweight-at-5 child – and what happens between 5 and 8? What is the predictive ability for excess growth on obesity (at an individual level)?

Our response: These are very interesting questions but not included in the scopes of our study. Nevertheless, excess infant growth is an established risk factor for the development of childhood obesity, as it is described in the introduction.

• Pre-term babies (if included): these would not necessarily be SGA, right? How did you treat their age (actual age or allowing for extra period)?

Our response: In the previous study by Sengpiel et al[3], within the MoBa study, caffeine intake, defined in the same way as in our study, was not associated with reduced length of gestation or preterm delivery. Hence we had no evidence to apply any restrictions on preterm children (i.e exclude, stratify). Nevertheless, we have included gestational age (as exact age determined by second-trimester ultrasound in 98.3% of pregnancies and based on the last menstrual period in the remaining cases) as a-priori confounder in all our models. We did not correct the age of preterm babies for an extra-period.

Reviewer: 3. Reviewer Name: Darren Greenwood

Institution and Country: University of Leeds, UK.

Please state any competing interests: I have a PhD student who is accessing MoBa data.

Please leave your comments for the authors below

bmjopen-2017-018895

Comments for the authors:

General comments:

1. MoBa is a superb resource from a highly esteemed research team.

2. The Generation R study demonstrated greater maternal caffeine intake to be associated with higher body mass index from 6 months onwards, slightly dampening the authors' claims to be the first to investigate this in infancy rather than childhood. However, the work is substantial, certainly still novel enough for publication, of very high standard and from a much larger cohort than Generation R.

Major points:

3. The statistical methods appear very sound. The growth models are good. The consideration of confounding is good, including what not to adjust for postnatally, though important residual confounding may remain, as the authors acknowledge. The consideration of incomplete data is good.

Our response: We would like to thank the reviewer for his general feedback on our manuscript.

4. The odds ratios are small, but obesity is common. It would be good to quantify this more clearly when interpreting the results. We always need to say "relative to what?" when we quote relative risks, etc. from cohort studies.

Our response: It is not very clear to us whether this comment refers to the difference between OR and RR or about the etiological fraction. Here, we have estimated OR . The interpretation of OR>1 is that the exposed have higher odds to develop the disease than the non-exposed. For example, the OR of excess growth in infancy of 1.66 is interpreted as follows: the odds of children exposed to very high caffeine to experience excess growth is 1.66 times higher than the odds of low exposed children. Indeed, this is more difficult to interpret than RR. As the OR are small and the outcomes not very prevalent, the ORs are probably not so different compared to RR but we would prefer not over-interpreting the associations.

Alternatively and referring to the etiological interpretation of the reviewer's comment, if the effect was causal, as the number of highly exposed women is high, an intervention could prevent quite a lot of children of being overweight, even if the effect is small. In addition, given the measurement error in exposure assessment that generally tends to decrease the effect size, the measured association is likely underestimated compared to the real one. In order to put the estimated OR in perspective we have listed previously reported effect estimates on maternal smoking in pregnancy, a well-known risk factor for childhood obesity, and shown that our effect estimates of caffeine are in the range of the reported effect estimates for maternal smoking. Unfortunately, this effect estimate has not been published for the MoBa study:

Reference	Study population	Prevalence of outcome using IOTF cut-offs	OR for maternal smoking during pregnancy
4	Australian 6-year olds	17.9% overweight 10.1% obese	1.40 1.85
5	German 6-year olds	13.7% overweight 3.9% obese	1.5 2.2
6	Dutch preschoolers	18.5% overweight (including obesity)	1.33
7			

	Greek 11-year olds	41.9%	1.80
8			
	12 studies meta-analysis		1.33 for overweight 1.60 for obesity

5. Maternal smoking is an important confounder adjusted for in the modelling. However, it is an unusually important confounder, very strongly associated with maternal caffeine intake, and strongly associated with accelerated catch-up infant growth. It is therefore important that this is measured very precisely. Self-reported measures may not be precise enough. Categorisation into ever/ never is even worse and loses any dose-response with the confounder. With odds ratios of the magnitude reported in the manuscript, it is well within the range for some of this to be residual confounding by smoking. Smoking may also keep maternal pre-pregnancy BMI artificially low, so not fully adjusting for maternal BMI. Sensitivity analysis amongst the self-reported never smokers would help reduce some of these doubts (for supplemental material).

Our response: We agree with the reviewer that smoking during pregnancy is an important confounder and we have reported that the higher the maternal caffeine intake the more likely the mother was smoking during pregnancy. In the main manuscript, we have excluded women who reported smoking during pregnancy and the results were similar.

Nevertheless, we agree that misreporting smoking in pregnancy is common as women identify it as a bad habit. As suggested by the reviewer, we have repeated our analysis by restricting to women who answered "no" in the question "have you ever smoke?" in the baseline questionnaire around mid-pregnancy (never smoke: 26,040, 51% and ever smoke: 24,664, 49%). Almost all women who never smoke were nonsmokers during pregnancy (99.97%). After restriction to never smokers, our results are similar but have attenuated, with lower OR and wider confidence intervals, especially for the very high caffeine intakes (see table below).

	Average OR (95%CI)	High OR (95%CI)	Very high OR (95%CI)
Excess infant growth	1.16 (1.07,1.25)	1.22 (1.02,1.48)	1.22 (0.84,1.79)
Overweight			
3 years	1.01 (0.93,1.10)	1.38 (1.15,1.66)	1.20 (0.83,1.73)
5 years	0.98 (0.92,1,07)	1.19 (1.02,1.41)	1.19 (0.86,1.62)
8 years	0.89 (0.77,1.01)	1.25 (0.94,1.64)	0.70 (0.35,1.37)

[a]Excess growth is defined as a WHO weight-for-age z-score difference >0.67 between birth and age 12 months.

[b] Overweight and/or obesity, according to International Obesity Task Force definition.

Models adjusted for maternal age, parity, parental education, pre-pregnancy BMI, total energy intake, nausea and/or vomiting during pregnancy, paternal BMI, paternal smoking during pregnancy, gestational age and gender.

As expected the caffeine consumption in the never smokers is different than in the total population; with fewer women with high (4.4% vs. 7%) and very high (1.1% vs. 3.2%) caffeine intakes. This might reduce the statistical power of detecting statistically significant associations in the very high levels of exposure. Nevertheless, similar positive trends were observed and from average to high intakes, the dose response associations persisted, for all outcomes. As we have many supplemental analyses and we have already presented the associations after excluding smokers in pregnancy, we have not included this analysis in our manuscript.

6. My main concern is about whether it's just that small babies experience catch-up growth, and this really hasn't got anything more to do with maternal caffeine intake other than it's a predictor of having a small baby. And if so, what does that really add, because we already know that? It appears to be an association without much of an established mechanism related to caffeine as such. Is it caffeine specifically that drives accelerated growth, or is it just that maternal caffeine is associated with smaller babies and all babies born small are more likely to experience accelerated growth in infancy, regardless of maternal caffeine intake? Can we look at this in the caffeine non-consumers by way of sensitivity analysis (possibly for supplemental material)?

Our response: We have tried to assess the effect of caffeine on infant growth independently of fetal growth restriction, by excluding SGA infants and also by adjusting for birth weight; and similar results were obtained. We also tested the interaction between birth weight and caffeine intake to study whether the association between caffeine intake and postnatal growth could differ according to the size at birth. It is not clear to us how the association between fetal growth restriction and accelerated infant growth in non-caffeine exposed children (n=56), would be more informative, given the aim of our study. That is not to identify determinants of accelerated growth in infancy, but focus on a factor that has been reported to have a negative association on fetal growth, that is caffeine.

It has been well documented that growth restricted neonates have an increased risk for accelerated infant growth in infancy. Nevertheless, recent research shows that some perinatal factors can also have a direct effect on postnatal growth, independent of effects on fetal growth [9-11]. Such factors were maternal and paternal body size, smoking during pregnancy, socioeconomic status and length of gestation. And this seems also true for caffeine intake. This final argument has now been clarified in the revised manuscript (page 14-15, lines 397-404).

7. What about paternal caffeine intake during the pregnancy – is this available? This should have no association with birthweight, conditioning on the confounders. So adjusting for it in the model should not change the estimate for maternal caffeine consumption, providing all potential confounders for maternal caffeine have been included. Is there some form of sensitivity analysis that could be conducted using this? It's a bit tenuous, I know, and on reflection the authors may legitimately argue that this is a poor suggestion! But if there were residual confounding from smoking or unmeasurable socio-economic factors, and these were partially common to both partners, this might tease some of that out (possibly for supplemental material).

Our response: This is an interesting point from the reviewer and a great suggestion. As there are no reports of a relationship between paternal caffeine consumption and the child's growth we have performed a negative control analysis as [12 13], using paternal caffeine the exposure.

We have calculated the caffeine intake of the father using the caffeine concentrations and serving sizes as used for the mother's calculations (Supplementary Table 1) for 5 food items: filtered coffee, boiled coffee, espresso coffee, caffeinated soft-drink with sugar or artificially sweetened. Only 16,455 (32%) fathers had available information. Their median intake was 193mg/day (p5-p95: 0-493mg/day), with caffeine from coffee being the main contributor (median: 187 mg/day). Fathers were consuming statistical significantly more caffeine than their partners ($p<0.001$ for Wilcoxon matched-pairs signed-ranks test). The spearman correlation coefficient between maternal and paternal caffeine intakes was 0.15 (p-value<0.0001). However, paternal intake was increasing by increasing levels of maternal intake and 45% of mothers with very high intake were with partners in the highest quartile of caffeine intake (see table below).

Table. Paternal caffeine intake by maternal caffeine intake, during pregnancy (n=16,455).

	Maternal daily caffeine intake			
	Low (<50mg)	Average (50-199mg)	High (200-299mg)	Very high (≥300mg)
Paternal daily caffeine intake				
Continuous (median, IQR)	178 (255)	208 (246)	249 (214)	285 (193)
In quartiles (n, %)				
Q1 (<81mg)	2,253 (29%)	1,504 (20%)	168 (15%)	44 (14%)
Q2 (86-188 mg)	1,605 (21%)	1,495 (20%)	202 (18%)	51 (16%)
Q3 (193-351 mg)	1,950 (26%)	2,186 (30%)	356 (33%)	82 (25%)
Q4 (>355 mg)	1,832 (24%)	2,211 (30%)	371 (34%)	145 (45%)

We have explored the association between maternal caffeine intake without and with adjustment for paternal caffeine intake and the same for paternal intake. All models are adjusted for the same confounders as in the main analysis. We have explored the associations with excess infant growth (n=12,289) and overweight at 3 years (n=16,455). The associations are presented graphically in Figures 1 and 2 (see below).

For both the risk of excess infant growth and overweight at 3 years, the association with maternal caffeine intake changed negligibly after adjusting for paternal intake. Hence, the association between maternal caffeine intake and postnatal growth was not strongly biased by unmeasured confounding shared by the mother and the father

On the other hand, by using paternal caffeine intake as a negative control, the trend of the association with child's growth was similar to that of maternal caffeine intake, while the effect estimate was much lower. Hence, unmeasured confounding due to shared family environment could have biased our analysis on maternal caffeine intake, but probably only a little. This information has now been added in the revised manuscript in different sections: methods (page 9, lines 238-246), results (page 10, lines 268-274 & page 12, lines 321-329), discussion (page 13, lines 346-350), Supplementary Table 4 and Supplementary Figures 2 & 3.

Figure 1. Association between maternal and paternal caffeine intake during pregnancy and excess infant growth.

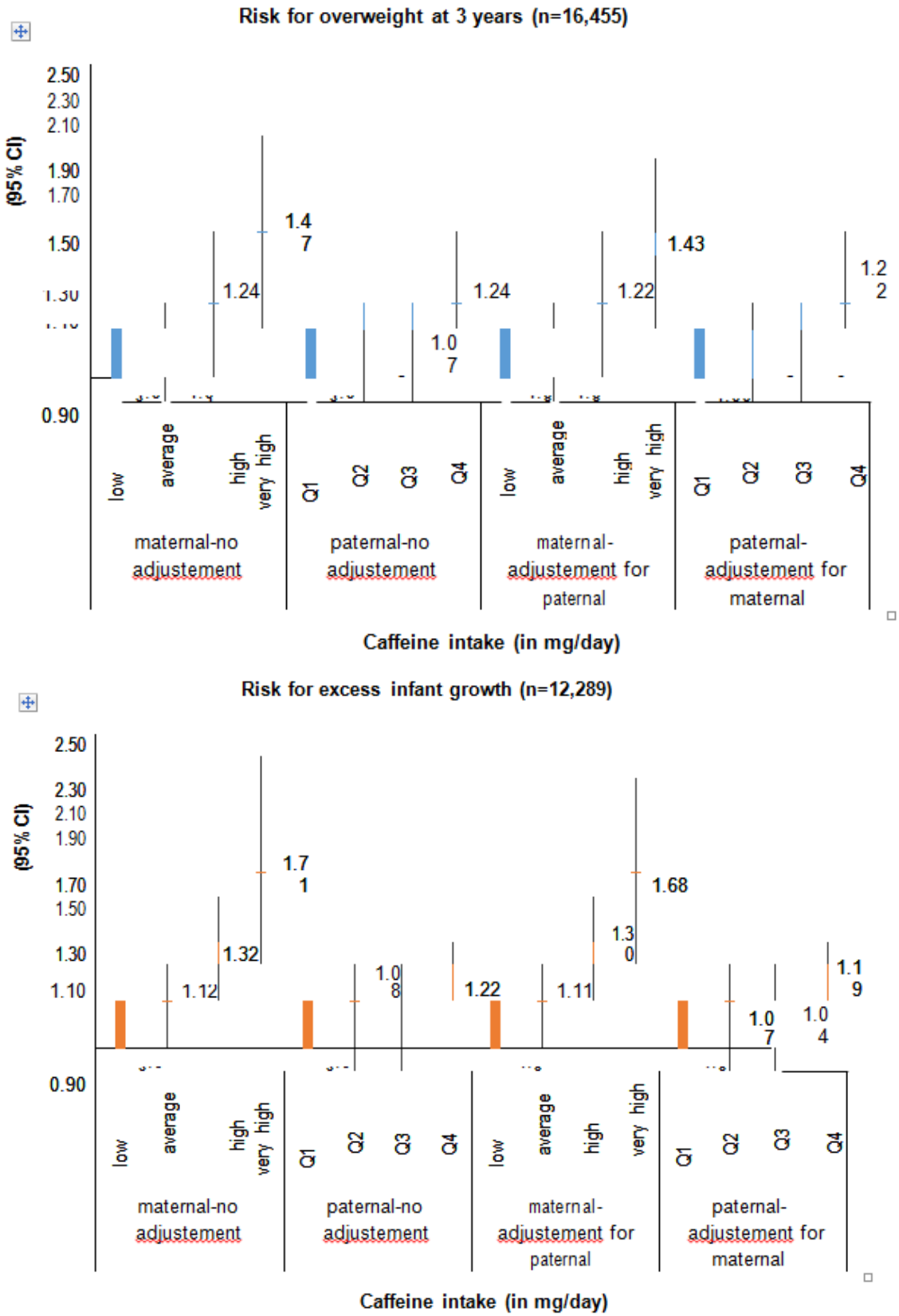


Figure 2. Association between maternal and paternal caffeine intake during pregnancy and overweight at 3 years.

Minor points:

8. Reference 4 could be expanded to cite some of those other authorities from a range of countries the authors refer to.

Our response: In the report of assessment of caffeine by the Norwegian Food Safety Authority, there is a detailed summary of previous risk assessments (section 2.1.1 and 2.1.2). Hence, we have used this report as a summarized reference of previous recommendations for caffeine consumption during pregnancy). We have now added the EFSA Scientific Opinion, published in 2015 (page 4, line 80).

9. Page 5, line 40. Presumably the mother completes the food frequency questionnaire, not the infant.

Our response: This sentence has now been revised by adding the phrase “pregnant women” (page 5, line 120).

10. Page 6, line 21 onwards. Presumably some of these units are mg/day.

Our response: We have now revised these units to mg/day (page 6, lines 139,140; page 11, line 303). We have not changed the units in the Tables as the top of the table is described that this is daily caffeine intake.

References

1. Maternal caffeine intake during pregnancy and risk of fetal growth restriction: a large prospective observational study. *BMJ* 2008;337:a2332. doi: 10.1136/bmj.a2332 [published Online First: 2008/11/05]

2. Norwegian Mother and Child Study: Norwegian Institute of Public Health; [Available from: http://www.fhi.no/eway/default.aspx?pid=238&trg=MainArea_5811&MainArea_5811=5895:0:15,3046:1:0:0:::0:02010.

3. Sengpiel V, Elind E, Bacelis J, et al. Maternal caffeine intake during pregnancy is associated with birth weight but not with gestational length: results from a large prospective observational cohort study. *BMC Med* 2013;11:42. doi: 10.1186/1741-7015-11-42

4. Gopinath B, Baur LA, Burlutsky G, et al. Socio-economic, familial and perinatal factors associated with obesity in Sydney schoolchildren. *J Paediatr Child Health* 2012;48(1):44-51. doi:

10.1111/j.1440-1754.2011.02181.x

5. von Kries R, Bolte G, Baghi L, et al. Parental smoking and childhood obesity--is maternal smoking in pregnancy the critical exposure? *International journal of epidemiology* 2008;37(1):210-6. doi: 10.1093/ije/dym239

6. Heppe DH, Kiefte-de Jong JC, Durmus B, et al. Parental, fetal, and infant risk factors for preschool overweight: the Generation R Study. *Pediatric research* 2013;73(1):120-7. doi: 10.1038/pr.2012.145

7. Magriplis E, Farajian P, Panagiotakos DB, et al. Maternal smoking and risk of obesity in school children: Investigating early life theory from the GRECO study. *Prev Med Rep* 2017;8:177-82. doi: 10.1016/j.pmedr.2017.10.001

8. Riedel C, Schonberger K, Yang S, et al. Parental smoking and childhood obesity: higher effect estimates for maternal smoking in pregnancy compared with paternal smoking--a meta-analysis. *International journal of epidemiology* 2014;43(5):1593-606. doi: 10.1093/ije/dyu150

9. Liu JX, Xu X, Liu JH, et al. Association of maternal gestational weight gain with their offspring's anthropometric outcomes at late infancy and 6 years old: mediating roles of birth weight and breastfeeding duration. *International journal of obesity* 2017 doi: 10.1038/ijo.2017.183

10. Hindmarsh PC, Geary MP, Rodeck CH, et al. Factors predicting ante- and postnatal growth. *Pediatric research* 2008;63(1):99-102. doi: 10.1203/PDR.0b013e31815b8e8f

11. Morgen CS, Angquist L, Baker JL, et al. Prenatal risk factors influencing childhood BMI and overweight independent of birth weight and infancy BMI: a path analysis within the Danish National Birth Cohort. *International journal of obesity* 2017 doi: 10.1038/ijo.2017.217

12. Richmond RC, Al-Amin A, Smith GD, et al. Approaches for drawing causal inferences from epidemiological birth cohorts: a review. *Early Hum Dev* 2014;90(11):769-80. doi:

10.1016/j.earlhumdev.2014.08.023

13. Brew BK, Gong T, Williams DM, et al. Using fathers as a negative control exposure to test the Developmental Origins of Health and Disease Hypothesis: A case study on maternal distress and offspring asthma using Swedish register data. *Scandinavian journal of public health* 2017;45(17_suppl):36-40. doi: 10.1177/1403494817702324

VERSION 2 – REVIEW

REVIEWER	Darren Greenwood University of Leeds, UK
REVIEW RETURNED	01-Dec-2017

GENERAL COMMENTS	The authors have addressed all my points in great detail and I am satisfied with all responses. I reiterate that the statistical methods are all done very well. All that remains is for me to clarify one earlier (minor) point I did not make clearly and apologise for missing some of the extensive sensitivity analysis they had already done: 4. Sorry if my comment on quantifying the underlying risk of the outcome was unclear. I was just referring to a doubling in risk of a common outcome being more important than a doubling in the risk of a rare outcome. So I was just suggesting that the % of children who are overweight or obese in the wider population be clearly referred to when interpreting the estimates. That was all I meant. I leave it to the authors' discretion. I was not really thinking of measurement error here, but the authors should remember that it's only non-differential measurement error that tends to attenuate the estimates. For self-reported exposure

	such as diet (and any self-reported covariates such as amount smoked) it is unlikely to be non-differential, so harder to predict the direction of measurement error bias. But no changes needed. 5. Thank you. I had simply missed the sensitivity analyses excluding smokers.

REVIEWER	Sonja Wehberg Center for Clinical Epidemiology, Odense University Hospital, and Research Unit of Clinical Epidemiology, Institute of Clinical Research, University of Southern Denmark, Odense, Denmark
REVIEW RETURNED	22-Dec-2017

GENERAL COMMENTS	The manuscript has been revised according to the recommendations, and you addressed all issues in your detailed answer, thank you. I am still a bit puzzled about employing growth curve models, and afterwards model random intercepts for child, for children who only have one actual measurement, but I accept that you did not want to exclude them. I have a few minor points: 1. Supplementary table 1: It would be helpful to have a column "target age in months". Furthermore, what N is valid for the caffeine intake levels (last 4 columns)? 2. Methods page 6: in your reply, you state that an actual measurement were set to missing, when the difference to the model prediction was larger than 5SD - in the text it says 3SD 3. Discussion, page 13 line 360: again, if you eliminate measurement pairs which show differences (as above), you expect a high correlation. I suggest you remove this sentence. 4. Discussion, page 13, line 365: I believe the reference should be to supplementary table1?

VERSION 2 – AUTHOR RESPONSE

Letter of response to reviewers'' comments
Manuscript ID: bmjopen-2017-018895.R1
Title: "Maternal caffeine intake during pregnancy is associated with excess growth in infancy and overweight in childhood: results from a large prospective cohort study"

On behalf of the co-authors, I would like to thank the reviewers for their effort and their valuable suggestions to our work. Please note that the page and line numbers are referred to the revised document with track changes.

Reviewer(s)' Comments to Author:
Reviewer: 3
Reviewer Name: Darren Greenwood
Institution and Country: University of Leeds, UK
Please state any competing interests or state 'None declared': I have a PhD student currently accessing MoBa data.

Please leave your comments for the authors below:
The authors have addressed all my points in great detail and I am satisfied with all responses. I reiterate that the statistical methods are all done very well.
All that remains is for me to clarify one earlier (minor) point I did not make clearly and apologise for missing some of the extensive sensitivity analysis they had already done:

4. Sorry if my comment on quantifying the underlying risk of the outcome was unclear. I was just referring to a doubling in risk of a common outcome being more important than a doubling in the risk of a rare outcome. So I was just suggesting that the % of children who are overweight or obese in the wider population be clearly referred to when interpreting the estimates. That was all I meant. I leave it to the authors' discretion.
I was not really thinking of measurement error here, but the authors should remember that it's only non-differential measurement error that tends to attenuate the estimates. For self-reported exposure such as diet (and any self-reported covariates such as amount smoked) it is unlikely to be non-differential, so harder to predict the direction of measurement error bias. But no changes needed.
Our response: we would like to thank the reviewer for the clarification. We have now included the following sentence in the discussion (page 14, lines 402-405):
"Given that overweight in childhood is not a rare condition and the number of children highly exposed to caffeine during pregnancy is large, even a small increase in the risk of overweight due to caffeine can result into a large proportion of children becoming overweight, assuming that the effect was causal."

5. Thank you. I had simply missed the sensitivity analyses excluding smokers.

Reviewer: 2
Reviewer Name: Sonja Wehberg
Institution and Country: Center for Clinical Epidemiology, Odense University Hospital, and Research Unit of Clinical Epidemiology, Institute of Clinical Research, University of Southern Denmark, Odense, Denmark
Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below:

The manuscript has been revised according to the recommendations, and you addressed all issues in your detailed answer, thank you.
I am still a bit puzzled about employing growth curve models, and afterwards model random intercepts for child, for children who only have one actual measurement, but I accept that you did not want to exclude them.

I have a few minor points:
1. Supplementary table 1: It would be helpful to have a column "target age in months". Furthermore, what N is valid for the caffeine intake levels (last 4 columns)?
Our response: we have now revised the Supplementary Table 1 with a column called "target age", with the age of the child when weight and height measurements were reported, as phrased in the questionnaires. Regarding the N for the caffeine intake levels, this has now been added in the table and it is the same as the N for weight.

2. Methods page 6: in your reply, you state that an actual measurement were set to missing, when the difference to the model prediction was larger than 5SD - in the text it says 3SD
Our response: we apologize for this mistake. The cut-off was set at 5SD and now this is corrected in the manuscript (page 6, line 153).

3. Discussion, page 13 line 360: again, if you eliminate measurement pairs which show differences (as above), you expect a high correlation. I suggest you remove this sentence.
Our response: we have now deleted the respective sentence (page 12, lines 331-332).

4. Discussion, page 13, line 365: I believe the reference should be to supplementary table1?
Our response: yes, we have now revised the reference to the correct table (page 12, line 336 & line 339).