Dear Editor

We thank the Reviewers for their thorough and relevant comments on our paper "Educational inequalities in cardiovascular risk factors: The impact of early life family factors in 228,346 Norwegian siblings" (PUBH-D-16-00132). We have addressed the comments point by point below. Changes in the paper are marked by tracked changes in the resubmitted version of the manuscript.

Reviewer reports:

Reviewer #1: The relationship between family factors of siblings aged 40-45 year olds and CVD risk factors were assessed in 228,346 individuals. The fact that you have sibling groups with differing educational level between two or more siblings is a strength of your study. Another strength is the sex-stratified analyses. I believe that you could be clearer with your results / summary. I look forward to reading the revision.

MAJOR
From your abstract I was expecting that a range of family factors such as number of siblings, housing location (city vs country), etc were being assessed. Whereas you are assessing the similarities between siblings as a measure of family factors that impact upon CVD. Please consider changing the abstract, and elsewhere in manuscript: Instead of "As a more robust approach we investigated to what extent family factors shared by siblings impact on inequalities in CVD risk factors in adulthood using discordant sibling design." -> "As a more robust approach we investigated to what extent family factors shared by siblings share similarities impact on inequalities in CVD risk factors in adulthood using discordant sibling design." OR "we investigated to what extent being from the same family impacted upon CVD risk factors in adulthood using discordant sibling design." OR "we investigated to what extent family factors shared by siblings share similarities impact on inequalities in CVD risk factors in adulthood using discordant sibling design." In contrast, you explain it very well in the introduction, and it could definitely be clearer in the abstract.

Author response: We appreciate the Reviewers suggestion to specify our aim and to define the term “family factors” more clearly.

Changes made:

We have rephrased the last part of the title from “The impact of early life family factors” to “Impact of shared family factors”.

Abstract, Background section, the second sentence

“As a more robust approach we investigated to what extent family factors shared by siblings impact on inequalities in CVD risk factors in adulthood using discordant sibling design” was changed into:

“As a more robust approach we investigated the impact of shared family factors on the educational gradient in midlife CVD risk factors by assessing within-sibling similarity in the gradient using a discordant sibling design”

Abstract, Conclusion, the first sentence

“About one third of the educational gradients in modifiable CVD risk factors may be explained by confounding from early life family factors” was changed into:

“About one third of the educational gradients in modifiable CVD risk factors may be explained by factors that siblings share”.

I'm not sure how you should deal with risk factors that are within the SCORE. Eg. Accounting for age and smoking. Have you consulted a statistician?
Author response: We appreciate the Reviewer’s suggestion to clarify the calculation of the SCORE risk score. Our colleague and statistician Randi Selmer, acknowledged in the Acknowledgments section, has provided the syntax for the calculation of the SCORE risk score based on the algorithm presented in Appendix A in the paper by Conroy and co-workers (Eur Heart J. 2003, 987-1003), and chose the coefficients for CVD risk in a high risk population. We included the variables from our study population on age, current daily smoking, systolic blood pressure and serum total cholesterol.

Changes made:

In the Methods section, under Cardiovascular risk factors, the last sentences was changed from “Systematic COronary Risk Evaluation (SCORE) risk prediction score of 10-year CVD mortality was calculated based on age, sex, systolic blood pressure, total cholesterol and daily smoking” into “Systematic COronary Risk Evaluation (SCORE) risk prediction score of 10-year CVD mortality was calculated based on our variables; age, sex, systolic blood pressure, total cholesterol and current daily smoking”

In addition a sentence was added: “We applied the algorithm presented in Appendix A in the paper by Conroy et al. and chose the coefficients for CVD risk in a high risk population.”

MINOR

Location (country/city) missing from abstract

Author response: we appreciate the opportunity to report the study location.

Changes made:

Abstract, first sentence: “Health survey data” was changed into “Norwegian health survey data”

"known parental status" do you mean unreported? Or do the participants not know their parental details? Similarly "known length of education"

Author response: we appreciate the opportunity to specify the terms “known” and “unknown”.

Changes made:

In the Results section, under Population, the term “known” was replaced with “registered”, and the term “unknown” was replaced by “unregistered”.

"Women had numerically stronger cohort and within sibships educational gradients for blood pressure, total cholesterol and BMI, and weaker gradients for the SCORE risk score (which is weighted for sex) (Supplementary materials S1 Tables A and B)." -> compared to? If you are comparing men to women, you would need to undertake another test and report the values and statistical significance.
Author response: we have not undertaken a formal statistical test to compare the estimates in men and women, and we are glad to nuance these statements.

Changes made:

In the Results section, under Sensitivity analyses, the first two sentences were changed from “Sex-stratified analyses included 73,023 sisters in 33,602 sibships and 65,400 brothers in 30,209 sibships. Women had numerically stronger cohort and within sibships educational gradients for blood pressure, total cholesterol and BMI, and weaker gradients for the SCORE risk score (which is weighted for sex) (Supplementary materials S1 Tables A and B).” into “Sex-stratified analyses included 73,023 sisters in 33,602 sibships and 65,400 brothers in 30,209 sibships, and showed a similar pattern as the main results. We did not formally test the sex differences in the estimates; however the educational gradients for blood pressure, total cholesterol and BMI in women were numerically stronger than in men(Supplementary materials S1 Tables A and B)”

How much cross over is there between 1980-1999 and 1990-2003 cohorts? Why is this analysis important? Maybe there was less education variability in the 1980-1999 cohort than the 1990-2003?

Author response: The thought behind this sensitivity analysis was to make an attempt to study any period effects. In order not to split too many sibships we used overlapping time periods. The analyses is however not easily interpretable both because of the cross over and the fact that we have not performed a statistical test of the potential differences in estimates between the two periods. We suggest that we remove this sensitivity analysis.

Changes made:

In the methods section, under Sensitivity analysis the third last sentence in deleted: “The analyses were stratified into two overlapping time periods for year of examination; 1980-1999 and 1990-2003”

In the Methods section, under Sensitivity analyses, the second last sentence is removed: “The within sibship similarities tended to explain somewhat more of the educational gradient in CVD risk factors for siblings examined during 1980-1999 than for siblings examined in 1990-2003 (Supplementary materials S2 Table)”

In the Discussion section, third paragraph, fourth sentence, the phrase “stratifying the analyses by an early and a late health survey period” was deleted.

In the Supplementary materials, Table S2 was deleted and Table S3 was renamed Table S2.

"In the sub-population of only sibships discordant for educational attainment the cohort estimates were slightly weaker than the cohort estimates in the study population (Supplementary materials, S3 Table vs. Table 3).” -> You would need to undertake another test and report the values and statistical significance for this specific test between the sub-population and the study population. I can't see where this has been undertaken.
Author response: we have not undertaken a formal statistical test to compare the estimates in men and women, and we are glad to nuance this statements.

Changes made:

In the Results section, under Sensitivity analyses, last sentence “In the sub-population of only sibhips discordant for educational attainment the cohort estimates were slightly weaker than the cohort estimates in the study population (Supplementary materials, S3 Table vs. Table 3)” was changed into “Analysis performed in the sub-population of only sibhips discordant for educational attainment showed similar pattern as in the main analyses(Supplementary materials, S2 Table vs. Table 3)”

Discussion - why "early life family factors"? are they not just "family factors"?

Author response: We agree that the term “family factors” might be more appropriate than the term “early life family factors”

Changes made:

Title, last part: “Impact of early life family factors in 228,346 Norwegian siblings” was rephrased into “Impact of shared family factors in 228,346 Norwegian siblings”.

Abstract, conclusion section, first sentence: “confounding from early life family factors” was changed into: “confounding from family factors”

In the Introduction, the Discussion and in the Conclusion, the term “early life family factors” was replaced by the term “family factors”.

I am a little confused by your discussion summary , could you make your points clearer? There is no discussion of CVD risk?

Discussion out of order. Main results summary, discussion of main results in literature, strengths and weaknesses, conclusion

Author response: we appreciate the reviewer’s suggestion to clarify both the discussion summary points and the disposition of the Discussion section, and have structured the phrases into four parts (main results summary, strengths and limitations, comparison with previous literature and interpretation of the results) and have added introductory phases denoting the disposition of the paragraphs.

Changes made:

In the discussion section, third paragraph, the first sentence was changed from
“The within sibships analysis conditions on sibships discordant for education and CVD risk factors” into

“A potential weakness of the within sibships analysis is that it conditions on sibships discordant for education and CVD risk factors”

In the Discussion section, next last paragraph, the first sentence was deleted “The on average 30% confounding of the educational CVD risk factor gradients from factors shared by siblings is considerable, and include both the environment that siblings share;” and replaced by “The sibling similarities in the educational gradients in the modifiable midlife CVD risk factors, blood pressure, serum total cholesterol, smoking, heart rate and BMI, that we identify (Table 3), can be interpreted as a result of both the environment that siblings share; (…)”

In the Discussion section, the last paragraph was rewritten and moved into the end of the paragraph discussing strengths and limitations.

Here, the first sentence is changed from

“Adult height is a composite measure of genetic disposition, health and nutrition during the growing years” into

“Our analyses were strengthened by including adult height which is a composite measure of genetic disposition, health and nutrition during the growing years, (…)”

Here, the last two sentences was rewritten from

“Consequently we suggest that the educational gradient in height that is not explained by shared early life family factors might result from residual confounding from non-shared factors in early life. This allows us to gauge the ability of the design to capture unobserved early life confounding.” into

“Accordingly, shared family factors had the strongest impact on the educational gradient in height, of all the CVD risk factor gradients(Table 3). Our sibling design has thus the potential to capture unobserved shared confounding, as the educational gradient in height that is not explained by shared family factors might result from residual confounding from non-shared factors in early life, childhood and youth.”

In the Conclusion section a phrase is added to the last sentence from “This suggests that there is substantial scope for prevention of CVD that starts early” into

“This suggests that there is substantial scope for prevention of CVD that starts early and that childhood environment matters”

DISCRETIONARY
Please replace the word "inequalities" as can be confused with socio-economical inequalities? Author response: We agree that the term “inequalities” needs to be specified more clearly.

Changes made:

Title, first part: “Educational inequalities in cardiovascular risk factors” was rephrased into “The educational gradient in cardiovascular risk factors”.

In the Introduction section in the first and third paragraph, the term “inequality” was omitted, and rephrased into “the inverse associations between SEP and CVD”, “the higher prevalence of CVD risk factors by lower SEP” or “socioeconomic gradient”.

In the Introduction section in the third and last paragraph the term “educational inequality” was replaced by the term “educational gradient”.

"a larger proportion of the educational gradient in most of the CVD risk factors was explained by similarities between siblings in siblings born closer in time” -> Are you inferring that siblings born closer together have more similar education status or have more similar CVD risk factors? OR both - hence the education status is a key player in CVD risk?

Author response: We are grateful for the Reviewer’s suggestion to give a clearer description and interpretation of this finding.

Changes made:

In the Results section under sibling analyses the second paragraph was changed from “In a subsample of sibships with only two siblings, analyses stratified for age-difference within a sibling pair indicated that a larger proportion of the educational gradient in most of the CVD risk factors was explained by similarities between siblings in siblings born closer in time (Table 4)” into

“In a subsample of sibships with only two siblings, analyses stratified for age-difference within a sibling pair indicated that a relatively larger proportion of the educational gradient in most of the CVD risk factors was explained by sibling similarities for the sibling pairs born closer in time than for the sibling pairs with larger age-span (Table 4)”

In the Discussion section, second last paragraph, the last sentence “These early life family factors were weakened by a larger age-span between siblings (Table 4), supporting that siblings with longer age–span share the same proportion of genes, but to a lesser extent their childhood environment, and suggesting that a large contribution of the family effect seems to be early environment rather than genes” was changed into

“These shared family similarities were weakened by a larger age-span between siblings (Table 4). This is in agreement with siblings with longer age–span sharing the same proportion of genes,
but to a lesser extent their childhood environment than sibling pairs born closer in time, and suggests that family environment plays an important role in the socioeconomic gradient in CVD”.

**TYPOS**

Author response: We very much appreciate the Reviewer’s suggestions to improve the standard of our English language and to remove typing errors.

Like -→ this is colloquial English

Changes made:

In the Introduction section, first paragraph, the last sentence was changed from

“Socioeconomic inequalities in CVD represent a potential for disease prevention, and it is of great policy interest to determine which periods during life course are important for development of cardiovascular disease risk factors like elevated blood pressure, disadvantageous serum lipid profile, obesity, tobacco use and physical inactivity.” into “Socioeconomic inequalities in CVD represent a potential for disease prevention, and it is of great policy interest to determine which periods during life course are important for development of the cardiovascular disease risk factors; elevated blood pressure, disadvantageous serum lipid profile, obesity, tobacco use and physical inactivity.”

"the educational gradient in height may”?

Changes made:

In the Introduction section, second paragraph, the last sentence “the educational gradient in height may serve as an instrumental variable for early life environment” was changed into “In this context, sibling similarity in the educational gradient in height might serve as a strong indicator of shared family environment.”

"with a with a model not using this approach"

Changes made:

In the introduction section, third paragraph, the second sentence the typing error was corrected into “with a model not using this approach”.

"Our results did was not altered"

Changes made:
In the Discussion section, third paragraph, sixth sentence, the typing errors were corrected: “Our results were not altered…”

"The on average 30%"

Changes made:

In the Discussion section, next last paragraph, the first sentence was deleted. “The on average 30% confounding of the educational CVD risk factor gradients from factors shared by siblings is considerable”

Reviewer #2: An interesting paper.

However, I recommend the authors to further detail their statistics and the CV risk factors assessment methods.

Author response: We agree that both the assessment methods for CVD risk factors and the statistical methods could be further detailed.

Changes made:

In the Methods section, under Cardiovascular risk factors, after the second sentence: “Smoking status and cigarette pack years were collapsed into a graded variable” we have added: “1) never smoker, 2) past smoker with < 20 pack years, 3) past smoker with ≥ 20 pack years, 4) current smoker with < 20 pack years, 5) current smoker with ≥ 20 pack years”

In the Methods section, under Cardiovascular risk factors, after the third sentence: “and the second of two measurements defined systolic and diastolic blood pressure. Later, the average of the last two available automatic oscillometric measures (Dinamap, Criticon, Tampa, USA) defined blood pressure.”

In the Method section, under cardiovascular risk factors, a phrase was added to the fifth sentence “Height and weight were measured and BMI (kg/m2) was calculated”

In the Methods section, under Cardiovascular risk factors, the last sentences was changed from “Systematic COronary Risk Evaluation (SCORE) risk prediction score of 10-year CVD mortality was calculated based on age, sex, systolic blood pressure, total cholesterol and daily smoking” into “Systematic COronary Risk Evaluation (SCORE) risk prediction score of 10-year CVD mortality was calculated based on our variables; age, sex, systolic blood pressure, total cholesterol and current daily smoking”

In addition a sentence was added: “We applied the algorithm presented in Appendix A in the paper by Conroy et al. and chose the coefficients for CVD risk in a high risk population.”

In the Methods section, under Statistical analyses, at the end of the first paragraph we have added a new paragraph with explanation of the interpretation of the multilevel analyses:
“Evaluation of the dissimilarity between the cohort and within sibship estimates helps interpret the role of shared family factors. If the between- and within-associations are equal, it indicates that the unobserved family-specific factors in the within sibships analyses are not important. A weaker within sibships association than cohort association may indicate that these unmeasured family factors confound parts of the association. A stronger within sibships association may indicate that these unmeasured factors have obscured the association. One example could be; among 4 siblings with different attained educational level, all overweight with a slightly lower BMI for the ones with higher educational level, there is a weaker within sibships educational gradient in BMI than among unrelated individuals in a cohort. If numerous sibships had such weaker gradients, the attenuation of the educational BMI gradient from the cohort analyses to the within sibships analyses would suggest that unmeasured family factors confound the association between educational level and BMI.”