nature portfolio

Peer Review File

Developmental origins of exceptional health and survival: A four-generation family cohort study

Corresponding Author: Mr Matthew Keys

This manuscript has been previously reviewed at another journal. This document only contains information relating to versions considered at Nature Communications.

Parts of this Peer Review File have been redacted as indicated to remove third-party material.

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #2

(Remarks to the Author)
I have previously reviewed (R3) this work in another journal, which allows me to focus on my prior comments.

The authors have generally addressed my previous concerns through careful revisions. Their response is detailed and well-argued, and the changes made to the manuscript are relevant. However, particular areas, such as the handling of socioeconomic and behavioral measures, the decision not to employ more robust validation techniques, and the continued discussion of genetic/epigenetic mechanisms, still present potential weaknesses. Here, I comment on some issues that remain.

While the authors made changes to the abstract, the absence of any explicit mention of limitations might still be a shortcoming, even if this was due to word count constraints.

Although justified, the decision not to incorporate more robust validation techniques remains a limitation. The authors argue that these techniques are more suited for prediction-focused studies rather than explanatory analyses like theirs. While this is a valid point, it's worth noting that even in explanatory research, cross-validation is definitely helpful to check for model stability and generalizability, particularly in large datasets where overfitting is a concern. Without additional validation techniques, the manuscript might be vulnerable to criticisms of overfitting, especially in the context of large sample sizes. The authors did perform checks for multicollinearity and provided risk differences, but these steps might not fully address concerns about the robustness of the models. After these suggested analyses, I would like to see the results remain the same before deciding on their robustness.

Using the term "mechanisms" in the context of this study is still problematic because it might imply a level of biological inference that the data does not support. The authors toned down their conclusions and removed strong claims about epigenetic mechanisms. However, the study still suggests that the observed health advantages could be compatible with heritable biological factors. While this hypothesis is reasonable, the lack of direct genetic or epigenetic data means these suggestions remain speculative.

While additional covariates were included, the discussion could have been more explicit about the potential remaining limitations due to unmeasured confounders, even after including new variables.

I was concerned about the limited scope of socioeconomic and behavioral indicators used in the study. The authors addressed this by including additional socioeconomic variables (e.g., parental income, marital status) and performing sensitivity analyses. However, even with these additions, the measures remain relatively simplistic. For instance, socioeconomic status is a multi-dimensional construct that includes factors like occupational prestige, profession, quality of education, wealth, and neighborhood effects, none of which are fully captured by the added variables. Furthermore, smoking as a binary measure fails to capture the intensity, duration, or quitting behaviors, which are critical for understanding its

impact on health outcomes.

The non-significant trends were justified by pointing to consistent patterns across outcomes and composites. While composite outcomes can provide a broader view, they can also mask important variations within individual components. Relying on patterns rather than statistically significant findings could lead to speculative conclusions. Over-reliance on trends that are not statistically significant could weaken the scientific rigor of the study. Readers might perceive this as an attempt to overstate the results, reducing the credibility of the findings.

Reviewer #3

(Remarks to the Author)
The authors have addressed my comments satisfactorily. I have no further comments.

Reviewer #4

(Remarks to the Author)
Thank you for the opportunity to review your manuscript titled "Developmental Origins of Exceptional Health and Survival: A Four-Generation Family Cohort Study." This study provides important insights into the familial aggregation of longevity and early life health outcomes, and I commend the authors for the rigorous approach in addressing a complex, multigenerational cohort.

Below, I outline my specific feedback, which I hope will help enhance the clarity and robustness of your manuscript.

General Comments:

The study is well-conceived and the use of Danish longitudinal registers is commendable, providing an impressive sample size and a rich source of data. The focus on early life health in longevity-enriched families adds significant value to the field of developmental origins of health and disease (DOHaD).

The primary finding of the manuscript is that descendants of longevity-enriched families (LEFs) in Denmark exhibit significant early life health and survival advantages, particularly in the third-generation (G3) grandchildren, compared to the general population. These advantages manifest as reduced infant mortality and lower risk of adverse neonatal outcomes, including preterm birth, small for gestational age, and neonatal respiratory disorders. However, this health advantage becomes attenuated in the fourth-generation (G4) great-grandchildren, suggesting a dilution of protective familial factors over successive generations. The study also finds that these advantages are independent of socioeconomic and behavioral factors, such as parental education and maternal smoking, and are observed in both maternal and paternal lines of transmission.

The observed reductions in preterm birth, small for gestational age, and neonatal respiratory disorders are more significant than neonatal and infant survival alone, as these outcomes demonstrate clear, robust, and relevant protective effects. These early-life advantages are well-established predictors of lower risks for chronic diseases and adverse health conditions later in life, aligning with the DOHaD framework.

There are several areas where the manuscript could be strengthened for clarity, interpretation, and broader relevance.

Specific Comments:

1. Significance of the Study
- The manuscript does a good job of highlighting the study's significance in relation to familial longevity and early health outcomes. However, the introduction would benefit from a more detailed discussion of how this study advances current understanding of both the DOHaD framework and the multigenerational transmission of health traits. Consider incorporating more context on how your study challenges or corroborates existing models of early life influence on adult health and survival.

2. Interpretation of Results: Attenuation Across Generations
- The observed attenuation of health and survival advantages in G4 compared to G3 descendants is an interesting finding. However, the manuscript does not fully explore the potential mechanisms behind this dilution effect. I recommend further discussion on whether this reflects a "catch-up" effect in the general population or a true loss of protective familial factors across generations.
- Additionally, consider exploring other potential environmental or behavioral shifts (e.g., healthcare improvements, lifestyle changes) or cohort effects that could be contributing to these observations, especially given the long study period.

3. Methodological Transparency and Reproducibility
- While the statistical approach is robust, I recommend expanding the discussion on the potential limitations of your methodological choices. Although the possibility of unmeasured confounding factors is acknowledged and addressed with an E-value analysis, further elaboration on how these confounders, especially in relation to socioeconomic status and parental education, could influence the results would enhance transparency. Additionally, while the E-value is a useful estimate of the potential impact, it's important to consider that multiple or numerous unmeasured confounders could contribute to an effect size as large as 3.07. Expanding on the plausibility and nature of these confounders, particularly

regarding socioeconomic factors and health behaviors, would further strengthen the interpretation of the findings.

4. Maternal Smoking and Behavioral Factors
- The reduced prevalence of maternal smoking in both G3 and G4 is an interesting observation. However, it would be valuable to explore this more deeply, as the behavioral mechanisms driving these differences remain unclear. Did you observe any significant differences in smoking cessation trends across pregnancy trimesters? Could maternal smoking be serving as a proxy for other unmeasured health behaviors? Additionally, given the incomplete data on maternal smoking for G3, it would be beneficial to acknowledge the limitations this presents when interpreting the findings and how it might influence the robustness of the conclusions.

5. Clarification of Figures and Tables
- The presentation of figures and tables is generally clear, but some areas would benefit from additional clarification. Specifically:
- In Figure 2, the Kaplan-Meier survival curves show divergence between G3 grandchildren and controls, but the G4 curves are less distinct. Adding more interpretation in the text to explain the implications of this divergence (or lack thereof) would be helpful.
- In Table 1, consider providing a brief explanation for why parity differs between G3 and G4, given that firstborn children are underrepresented in G3 due to exclusion criteria. This is mentioned briefly in the methods but could be emphasized in the results section for clearer interpretation.

6. Expansion on Sensitivity Analyses
- The sensitivity analyses conducted are a strength of the study. However, the manuscript could be improved by providing a more detailed discussion of the findings from these analyses. For instance, did the inclusion of unmatched children in sensitivity analyses materially affect your conclusions, and how did the analysis of more granular subgroups (e.g., very/extremely small for gestational age) contribute to your overall interpretation?

7. Potential Epigenetic and Genetic Mechanisms
- While you briefly mention the potential role of epigenetic mechanisms in multigenerational transmission of health traits, further exploration of this possibility could add valuable depth to your discussion. Although direct evidence may be lacking, discussing how epigenetic mechanisms might explain both the persistence and attenuation of observed health advantages would align your work with the growing body of research in the DOHaD framework. Acknowledging the speculative nature while considering future directions for research in this area would enhance the manuscript's relevance to current scientific debates. Recent evidence supporting both maternal and paternal transgenerational inheritance of health traits, such as altered DNA methylation and histone modifications, suggests that elaborating on this point is warranted.

8. Limitations and Future Research
- The limitations section would benefit from further elaboration, especially regarding the challenges of studying intergenerational health transmission over such an extended time frame. Consider discussing the potential for cohort effects or other time-sensitive variables (e.g., advancements in healthcare or neonatal care) that could influence the outcomes across generations.
- It would also be useful to briefly outline future research directions that could address these limitations, such as prospective studies of health outcomes in G4 great-grandchildren as they age.

Conclusion
In summary, this study offers a novel and valuable contribution to the literature on the developmental origins of health and the familial aggregation of longevity. I encourage you to consider the points raised in this review to further enhance the clarity, rigor, and relevance of the manuscript.

Version 1:

Reviewer comments:

Reviewer #2

(Remarks to the Author)
I have once again revised this manuscript. As I detailed in my earlier, more extensive critique of the first submission—and reiterated in the last one—I stated that even in explanatory research, cross-validation is required to assess model stability, generalizability, and overfitting, particularly in the context of large sample sizes. In this review, I only address my previous concerns and have not assessed the overall manuscript in other aspects.

Once again, the authors have refused to perform these analyses, arguing that "these methods are not able to produce accurate statistical inference (p-values, confidence intervals) on their estimated covariate parameters," which is simply not true. Their reluctance to conduct the requested analyses, combined with the flawed argument they have repeated, raises doubts about the quality of the work.

The authors argue that penalized regression models (e.g., LASSO) are not suitable for their explanatory research aims. However, I disagree entirely. my concern remains that cross-validation can help verify model stability even in explanatory models. The response includes a sensitivity analysis using penalized Cox models with cross-validation (via the glmnet

package in R). Still, it is unclear why they did not perform the requested analysis. This omission further increases doubts about the robustness of their results.

Other Points
Explicit mention of limitations in the abstract:

The response states that the wording was adjusted in the abstract to indicate limitations in measuring socioeconomic factors. However, I explicitly suggested adding a clear limitations statement, which has still not been done.

Over-reliance on non-significant trends: I was concerned about interpreting results that were not statistically significant but were used as evidence of trends. The response explains that composite outcomes were used to strengthen conclusions. However, they should report only global scores in such cases; otherwise, they must explicitly state that non-significant effects cannot be used as evidence. Such patterns should not be over-interpreted in the absence of statistical significance.

Reviewer #4

(Remarks to the Author)

Version 2:

Reviewer comments:

Reviewer #5

(Remarks to the Author)

Dear Reviewers,

Thank you for your excellent feedback on our study. A summary of the main changes to the manuscript, as guided by your comments, are as follows:

1. We have revised the discussion section to integrate additional comments on patterns of intergenerational transmission within our study and perspectives on potentially relevant epigenetic mechanisms.
2. We have revised the discussion section to improve the communication of relevant limitations of our study as emphasized by the reviewers, including those relating to maternal smoking and socioeconomic measurements.
3. We have provided a more thorough description of our numerous sensitivity analyses within the main manuscript, including reflections on how they affect the interpretation of our study.
4. We have provided additional analyses assessing the robustness of our findings within a prediction optimization framework to alleviate reviewer concerns regarding model overfitting.
5. We have revised the formatting of the manuscript in accordance with the requirements set out by *Nature Communications*.

In our responses below, we first present each review in full and then respond to all specific comments point-by-point. In all cases, where we refer to line numbers and references, they apply to the revised manuscript and supplementary material without tracked changes turned on.

Best regards,
Matthew Thomas Keys

Reviewer 2

I have previously reviewed (R3) this work in another journal, which allows me to focus on my prior comments.

The authors have generally addressed my previous concerns through careful revisions. Their response is detailed and well-argued, and the changes made to the manuscript are relevant. However, particular areas, such as the handling of socioeconomic and behavioral measures, the decision not to employ more robust validation techniques, and the continued discussion of genetic/epigenetic mechanisms, still present potential weaknesses. Here, I comment on some issues that remain.

While the authors made changes to the abstract, the absence of any explicit mention of limitations might still be a shortcoming, even if this was due to word count constraints.

Although justified, the decision not to incorporate more robust validation techniques remains a limitation. The authors argue that these techniques are more suited for prediction-focused studies rather than explanatory analyses like theirs. While this is a valid point, it's worth noting that even in explanatory research, cross-validation is definitely helpful to check for model stability and generalizability, particularly in large datasets where overfitting is a concern. Without additional validation techniques, the manuscript might be vulnerable to criticisms of overfitting, especially in the context of large sample sizes. The authors did perform checks for multicollinearity and provided risk differences, but these steps might not fully address concerns about the robustness of the models. After these suggested analyses, I would like to see the results remain the same before deciding on their robustness.

Using the term "mechanisms" in the context of this study is still problematic because it might imply a level of biological inference that the data does not support. The authors toned down their conclusions and removed strong claims about epigenetic mechanisms. However, the study still suggests that the observed health advantages could be compatible with heritable biological factors. While this hypothesis is reasonable, the lack of direct genetic or epigenetic data means these suggestions remain speculative.

While additional covariates were included, the discussion could have been more explicit about the potential remaining limitations due to unmeasured confounders, even after including new variables.

I was concerned about the limited scope of socioeconomic and behavioral indicators used in the study. The authors addressed this by including additional socioeconomic variables (e.g., parental income, marital status) and performing sensitivity analyses. However, even with these additions, the measures remain relatively simplistic. For instance, socioeconomic status is a multi-dimensional construct that includes factors like occupational prestige, profession, quality of education, wealth, and neighborhood effects, none of which are fully captured by the added variables. Furthermore, smoking as a binary measure fails to capture the intensity, duration, or quitting behaviors, which are critical for understanding its impact on health outcomes.

The non-significant trends were justified by pointing to consistent patterns across outcomes and composites. While composite outcomes can provide a broader view, they can also mask important variations within individual components. Relying on patterns rather than statistically significant findings could lead to speculative conclusions. Over-reliance on trends that are not statistically significant could weaken the scientific rigor of the study. Readers might perceive this as an attempt to overstate the results, reducing the credibility of the findings.

Dear Reviewer 2,

Thank you for your thorough review of our study. We respond to each of your comments point-by-point below.

1. *While the authors made changes to the abstract, the absence of any explicit mention of limitations might still be a shortcoming, even if this was due to word count constraints.*

Thank you for raising this point. We have modified the language used in the abstract to indicate that we measure a *select* number of key socioeconomic factors. However, all our estimates are highly robust to adjustment for the important factors we can measure (e.g. education, income and marital status), which minimizes the possibility of large attenuations of our estimates resulting from adjustment for other unobserved yet correlated factors. If, on the other hand, we had seen large attenuations in our estimates, then it would be likely that there may be residual confounding or meditation due to other related factors which were not measured. Therefore, we believe the current phrasing is appropriate given the limited word count. However, if the editor believes it would be worthwhile to expand the abstract beyond the normal word count for a more explicit discussion of these points, we can do so.

2. *Although justified, the decision not to incorporate more robust validation techniques remains a limitation. The authors argue that these techniques are more suited for prediction-focused studies rather than explanatory analyses like theirs. While this is a valid point, it's worth noting that even in explanatory research, cross-validation is definitely helpful to check for model stability and generalizability, particularly in large datasets where overfitting is a concern. Without additional validation techniques, the manuscript might be vulnerable to criticisms of overfitting, especially in the context of large sample sizes. The authors did perform checks for multicollinearity and provided risk differences, but these steps might not fully address concerns about the robustness of the models. After these suggested analyses, I would like to see the results remain the same before deciding on their robustness.*

Thank you for raising this point. We maintain our initial concerns over the rationale for the proposal to use penalized regression models in our setting, given the low risk of overfitting. Besides the concerns raised previously, it should be noted that most models of these varieties are not able to produce accurate statistical inference (p values, confidence intervals) on their estimated covariate parameters, with a few select exceptions due to recent advances. However, this has not yet been extended to the case of Cox proportional

hazard regression models, which were used in our study. In general, it would not be possible to alter our entire study design and explanatory research aims to use models which cannot produce accurate statistical inference.

However, we have undertaken several sensitivity analyses to assess whether our primary models on infant survival are affected by overfitting. We employed the *glmnet* package in R which allows estimation of Cox proportional hazard regression models for survival data by penalized partial likelihood, with 10-fold cross validation, using Harrel's C index as a performance measure. Since none of these models could be estimated from within our matched cohort design, we instead used the whole Danish population prior to the matching stage, where the matching criteria were instead employed as factor covariates. We report hazard ratios for the coefficient of interest (belonging to a longevity-enriched family) for all three models, and where appropriate, Harrel's C index and the values of λ employed:

I. Cox proportional hazard regression model estimated using the full Danish population (i.e. prior to matching) adjusted for sex and birth year only, no penalization or cross validation.
II. Cox proportional hazard regression model estimated using the full Danish population (i.e. prior to matching) adjusted for all primary demographic and socioeconomic covariates, with cross validation, but with the penalization parameter set to $\lambda = 0$ (equivalent to unpenalized Cox proportional hazard regression).
III. Cox proportional hazard regression model estimated using the full Danish population (i.e. prior to matching) adjusted for all primary demographic and socioeconomic covariates, with penalization parameter set to its optimal value $\lambda = \lambda^*$ as determined by 10-fold cross-validation.

The results of these analyses are reported in Item S3.14 of the Supplementary Appendix. Both our primary estimated effect and cross-validated C index and were stable across models 2 and 3, implying that our most saturated models are not susceptible to overfitting.

3. *Using the term "mechanisms" in the context of this study is still problematic because it might imply a level of biological inference that the data does not support. The authors toned down their conclusions and removed strong claims about epigenetic mechanisms. However, the study still suggests that the observed health advantages could be compatible with heritable biological factors. While this hypothesis is reasonable, the lack of direct genetic or epigenetic data means these suggestions remain speculative.*

Thank you for raising this point. We agree that using the term in this way could be misinterpreted. We now use the term "***epidemiological*** *mechanisms*" where relevant, which is an accurate portrayal of the scope of our study.

4. *While additional covariates were included, the discussion could have been more explicit about the potential remaining limitations due to unmeasured confounders, even after including new variables. I was concerned about the limited scope of socioeconomic and behavioral indicators used in the study. The authors addressed this by including additional socioeconomic variables*

(e.g., parental income, marital status) and performing sensitivity analyses. However, even with these additions, the measures remain relatively simplistic. For instance, socioeconomic status is a multi-dimensional construct that includes factors like occupational prestige, profession, quality of education, wealth, and neighborhood effects, none of which are fully captured by the added variables.

Thank you for raising this point. We have expanded some aspects of our discussion about the limitations of the socioeconomic variables (Lines 492-493 and 521-523).

5. *Furthermore, smoking as a binary measure fails to capture the intensity, duration, or quitting behaviors, which are critical for understanding its impact on health outcomes.*

Thank you for raising this excellent point. We have expanded the section on the limitations of maternal smoking, to incorporate these insights (Lines 519-521).

Reviewer 3

The authors have addressed my comments satisfactorily. I have no further comments.

Dear Reviewer 3,

Thank you for reviewing our revised manuscript, and we are glad could address your comments satisfactorily. We have made additional changes upon recommendations by other reviewers, which can be seen in our summary letter and specific responses.

Reviewer 4

Thank you for the opportunity to review your manuscript titled "Developmental Origins of Exceptional Health and Survival: A Four-Generation Family Cohort Study." This study provides important insights into the familial aggregation of longevity and early life health outcomes, and I commend the authors for the rigorous approach in addressing a complex, multigenerational cohort.

Below, I outline my specific feedback, which I hope will help enhance the clarity and robustness of your manuscript.

General Comments:

The study is well-conceived and the use of Danish longitudinal registers is commendable, providing an impressive sample size and a rich source of data. The focus on early life health in longevity-enriched families adds significant value to the field of developmental origins of health and disease (DOHaD).

The primary finding of the manuscript is that descendants of longevity-enriched families (LEFs) in Denmark exhibit significant early life health and survival advantages, particularly in the third-generation (G3) grandchildren, compared to the general population. These advantages manifest as reduced infant mortality and lower risk of adverse neonatal outcomes, including preterm birth, small for gestational age, and neonatal respiratory disorders. However, this health advantage becomes attenuated in the fourth-generation (G4) great-grandchildren, suggesting a dilution of protective familial factors over successive generations. The study also finds that these advantages are independent of socioeconomic and behavioral factors, such as parental education and maternal smoking, and are observed in both maternal and paternal lines of transmission.

The observed reductions in preterm birth, small for gestational age, and neonatal respiratory disorders are more significant than neonatal and infant survival alone, as these outcomes demonstrate clear, robust, and relevant protective effects. These early-life advantages are well-established predictors of lower risks for chronic diseases and adverse health conditions later in life, aligning with the DOHaD framework.

There are several areas where the manuscript could be strengthened for clarity, interpretation, and broader relevance.

Specific Comments:

1. Significance of the Study

- The manuscript does a good job of highlighting the study's significance in relation to familial longevity and early health outcomes. However, the introduction would benefit from a more detailed discussion of how this study advances current understanding of both the DOHaD framework and the multigenerational

transmission of health traits. Consider incorporating more context on how your study challenges or corroborates existing models of early life influence on adult health and survival.

2. Interpretation of Results: Attenuation Across Generations

- The observed attenuation of health and survival advantages in G4 compared to G3 descendants is an interesting finding. However, the manuscript does not fully explore the potential mechanisms behind this dilution effect. I recommend further discussion on whether this reflects a "catch-up" effect in the general population or a true loss of protective familial factors across generations.

- Additionally, consider exploring other potential environmental or behavioral shifts (e.g., healthcare improvements, lifestyle changes) or cohort effects that could be contributing to these observations, especially given the long study period.

3. Methodological Transparency and Reproducibility

- While the statistical approach is robust, I recommend expanding the discussion on the potential limitations of your methodological choices. Although the possibility of unmeasured confounding factors is acknowledged and addressed with an E-value analysis, further elaboration on how these confounders, especially in relation to socioeconomic status and parental education, could influence the results would enhance transparency. Additionally, while the E-value is a useful estimate of the potential impact, it's important to consider that multiple or numerous unmeasured confounders could contribute to an effect size as large as 3.07. Expanding on the plausibility and nature of these confounders, particularly regarding socioeconomic factors and health behaviors, would further strengthen the interpretation of the findings.

4. Maternal Smoking and Behavioral Factors

- The reduced prevalence of maternal smoking in both G3 and G4 is an interesting observation. However, it would be valuable to explore this more deeply, as the behavioral mechanisms driving these differences remain unclear. Did you observe any significant differences in smoking cessation trends across pregnancy trimesters? Could maternal smoking be serving as a proxy for other unmeasured health behaviors? Additionally, given the incomplete data on maternal smoking for G3, it would be beneficial to acknowledge the limitations this presents when interpreting the findings and how it might influence the robustness of the conclusions.

5. Clarification of Figures and Tables

- The presentation of figures and tables is generally clear, but some areas would benefit from additional clarification.

Specifically:

- In Figure 2, the Kaplan-Meier survival curves show divergence between G3 grandchildren and controls, but the G4 curves are less distinct. Adding more interpretation in the text to explain the implications of this divergence (or lack thereof) would be helpful.

- In Table 1, consider providing a brief explanation for why parity differs between G3 and G4, given that firstborn children are underrepresented in G3 due to exclusion criteria. This is mentioned briefly in the methods but could be emphasized in the results section for clearer interpretation.

6. Expansion on Sensitivity Analyses

- The sensitivity analyses conducted are a strength of the study. However, the manuscript could be improved by providing a more detailed discussion of the findings from these analyses. For instance, did the inclusion of unmatched children in sensitivity analyses materially affect your conclusions, and how did the analysis of more granular subgroups (e.g., very/extremely small for gestational age) contribute to your overall interpretation?

7. Potential Epigenetic and Genetic Mechanisms

- While you briefly mention the potential role of epigenetic mechanisms in multigenerational transmission of health traits, further exploration of this possibility could add valuable depth to your discussion. Although direct evidence may be lacking, discussing how epigenetic mechanisms might explain both the persistence and attenuation of observed health advantages would align your work with the growing body of research in the DOHaD framework. Acknowledging the speculative nature while considering future directions for research in this area would enhance the manuscript's relevance to current scientific debates. Recent evidence supporting both maternal and paternal transgenerational inheritance of health traits, such as altered DNA methylation and histone modifications, suggests that elaborating on this point is warranted.

8. Limitations and Future Research

- The limitations section would benefit from further elaboration, especially regarding the challenges of studying intergenerational health transmission over such an extended time frame. Consider discussing the potential for cohort effects or other time-sensitive variables (e.g., advancements in healthcare or neonatal care) that could influence the outcomes across generations.

- It would also be useful to briefly outline future research directions that could address these limitations, such as prospective studies of health outcomes in G4 great-grandchildren as they age.

Conclusion

In summary, this study offers a novel and valuable contribution to the literature on the developmental origins of health and the familial aggregation of longevity. I encourage you to consider the points raised in this review to further enhance the clarity, rigor, and relevance of the manuscript.

Dear Reviewer 4,

Thank you for your praise and thorough review of our study. We respond to each of your comments point-by-point below.

1. *The manuscript does a good job of highlighting the study's significance in relation to familial longevity and early health outcomes. However, the introduction would benefit from a more detailed discussion of how this study advances current understanding of both the DOHaD framework and the multigenerational transmission of health traits. Consider incorporating more context on how your study challenges or corroborates existing models of early life influence on adult health and survival.*

Thank you for raising this point. Early life health phenotypes have been associated with the risk of a range of chronic diseases in mid-life. However, many of these studies have been constrained in their perspective due to limited follow-up in birth cohorts that have sufficient data on early life health characteristics, and methodological constraints such as attrition. Some studies have taken a 'top-down' approach where late-life health and survival in extinct cohorts has been associated with intrauterine exposure to known environmental shocks in the 18th and 19th centuries, for example famine and economic recessions. However, there is little research on the relevance of these mechanisms to survival at advanced ages, outside of extreme environmental shocks in early life, particularly for phenotypes that aggregate in families. This is the gap that our study aimed to fill. In chorus with the wider literature, our findings suggest that the familial aggregation of exceptional health and survival longevity may have early developmental components and highlight the potential importance of these mechanisms governing exceptional longevity in general. These points are highlighted throughout the introduction section.

However, we think that a more detailed explanation of how our findings relate the DOHaD literature is better reserved for the discussion section of the manuscript. Here, we refer to observed advantages in outcomes that have been previously central to formulating the DOHaD hypothesis (e.g. birth weight), patterns in specific adult health outcomes in the parents from our previous research (e.g. mental and behavioral, cardiometabolic diseases), and the patterns of findings in our study that may implicate heritable biological factors in the transmission across generations. However, as raised by other reviews, certain aspects of this discussion are speculative or descriptive, e.g., the relevance of epigenetic mechanisms, or mediating relationships between early life health outcomes and late life phenotypes. For these reasons, we would argue that is better to limit discussion of these aspects in the introduction, to not mislead readers on the scope of our study. We hope you find this perspective and decision appropriate.

2. *The observed attenuation of health and survival advantages in G4 compared to G3 descendants is an interesting finding. However, the manuscript does not fully explore the potential mechanisms behind this dilution effect. I recommend further discussion on whether this reflects a "catch-up" effect in the general population or a true loss of protective familial factors across generations.*

Thank you for raising this excellent point. Due to the long follow-up period in our study, it is plausible that the lack of differences in our 4th generation analyses may reflect a 'catchup' effect of the general population, rather than a 'dilution' of the protective effect of LEF descendancy. Indeed, the balance of these two competing effects has important implications for the interpretation of our study. To address this issue, we performed our 'direct comparison' analysis (*Figure 3, right panel*) where survival in G3 descendants was compared to survival in G4 descendants directly in the periods of overlapping birth years between the two cohorts, with parametric adjustment for sex, maternal age, parity, birth year and birth season. This adjustment was necessary since the G3 descendants included in this subgroup were more likely to be from older mothers and not be a first child, whilst the G4 descendants included were more likely to be from younger mothers and be a first child. However, by adjusting for birth year as a factor, this essentially keeps all factors occurring in the background constant – including healthcare improvements, population wide behavior changes, and other cohort effects.

Our results from this analysis suggest a dilution of exceptional survival between G3 and G4 descendants of longevity-enriched siblings (HR = 0.32, 95% CI [0.16, 0.64]), even when keeping constant the trends in infant mortality and other secular and demographic variables over time. We also validated this methodological approach by assessing several negative control outcomes, where we were able to consistently retrieve expected associations (S3.1). From these analyses, we concluded that our findings likely reflect a true dilution of protective survival factors otherwise present in previous generations. These analyses are comprehensively described throughout the manuscript and in the relevant parts of the discussion (Lines 177-183, 205-214, 279-292, 397-402, 420-434). We have updated some sections in the discussion for clarity (Lines 425-426 and 427-429).

3. *Additionally, consider exploring other potential environmental or behavioral shifts (e.g., healthcare improvements, lifestyle changes) or cohort effects that could be contributing to these observations, especially given the long study period.*

Thank you for raising this excellent point. Please see our previous comment to point 2.

4. *While the statistical approach is robust, I recommend expanding the discussion on the potential limitations of your methodological choices. Although the possibility of unmeasured confounding factors is acknowledged and addressed with an E-value analysis, further elaboration on how these confounders, especially in relation to socioeconomic status and parental education, could influence the results would enhance transparency. Additionally, while the E-value is a useful estimate of the potential impact, it's important to consider that multiple or numerous unmeasured confounders could contribute to an effect size as large as 3.07. Expanding on the plausibility and nature of these confounders, particularly regarding socioeconomic factors and health behaviors, would further strengthen the interpretation of the findings.*

Thank you for raising this excellent point. Unmeasured confounding has been a concern raised by several reviewers in the interpretation of our findings. Although we measure only a *select* number of key socioeconomic and behavioral variables, our estimates are very robust to adjustment for these factors of

well-documented epidemiological importance, which reduces the possibility of large attenuations of our estimates resulting from adjustment for other unobserved yet correlated factors. If, on the other hand, we had seen large attenuations in our estimates, then it would be likely that there may be residual confounding/meditation due to other related factors which were not measured. We have updated the language used to state that our measures are but a limited set of potentially relevant socioeconomic and behavioral factors (Lines 492-493 and 521-523) and describe our perspective on the plausibility of unmeasured biases affecting our interpretations here (Lines 493-496).

5. *The reduced prevalence of maternal smoking in both G3 and G4 is an interesting observation. However, it would be valuable to explore this more deeply, as the behavioral mechanisms driving these differences remain unclear. Did you observe any significant differences in smoking cessation trends across pregnancy trimesters? Could maternal smoking be serving as a proxy for other unmeasured health behaviors? Additionally, given the incomplete data on maternal smoking for G3, it would be beneficial to acknowledge the limitations this presents when interpreting the findings and how it might influence the robustness of the conclusions.*

Thank you for raising this point. It is correct that maternal smoking will capture, by proxy, a range of other correlated but unmeasured behavioral and social factors. However, we cannot directly measure these due to the limitations of our register-based study. Similarly, data on quitting smoking throughout the trimesters of pregnancy were only available towards the end of our follow-up period and so could not be used for our study. We have previously addressed these limitations in the discussion section (Lines 514-521). However, given that this has been raised by another reviewer, we have revised our discussion of these points for clarification (Lines 519-521).

6. *In Figure 2, the Kaplan-Meier survival curves show divergence between G3 grandchildren and controls, but the G4 curves are less distinct. Adding more interpretation in the text to explain the implications of this divergence (or lack thereof) would be helpful.*

Thank you for raising this excellent point. We now describe the Kaplan-Meier survival curves and the implications of their trends in terms of inferences on neonatal versus infant survival (Lines 254-255, 256-257).

7. *In Table 1, consider providing a brief explanation for why parity differs between G3 and G4, given that firstborn children are underrepresented in G3 due to exclusion criteria. This is mentioned briefly in the methods but could be emphasized in the results section for clearer interpretation.*

Thank you for raising this concern. However, we already provide an explicit explanation for this point in the results section where we discuss Table 1 (Lines 231-235).

8. *The sensitivity analyses conducted are a strength of the study. However, the manuscript could be improved by providing a more detailed discussion of the findings from these analyses. For instance, did the inclusion of unmatched children in sensitivity analyses materially affect your*

conclusions, and how did the analysis of more granular subgroups (e.g., very/extremely small for gestational age) contribute to your overall interpretation?

Thank you for your praise on this point. We have added more content to the sensitivity analyses section of the manuscript, by expanding on the analyses we did and as well as their interpretation (Lines 375-377, 382-383, 384-385, 388-389, and 390-396).

9. *While you briefly mention the potential role of epigenetic mechanisms in multigenerational transmission of health traits, further exploration of this possibility could add valuable depth to your discussion. Although direct evidence may be lacking, discussing how epigenetic mechanisms might explain both the persistence and attenuation of observed health advantages would align your work with the growing body of research in the DOHaD framework. Acknowledging the speculative nature while considering future directions for research in this area would enhance the manuscript's relevance to current scientific debates. Recent evidence supporting both maternal and paternal transgenerational inheritance of health traits, such as altered DNA methylation and histone modifications, suggests that elaborating on this point is warranted.*

Thank you for raising this excellent point. In response to your comment, we have provided brief mentions to some of the recent developments in this area, including evidence for paternal and maternal transgenerational inheritance of health traits, and the potential mechanisms underlying these observations (Lines 478-479 and 481-482). We have also added references to several recent reviews and developments on this topic for interested readers (References 61-63, 65). However, the discussion of epigenetic factors has been a point of contention during peer-review of this manuscript and so we would also like to avoid putting too much emphasis on mechanisms that were not directly assessed in our study; this was raised by several other reviewers both previously and currently. We hope you find the balanced approach of briefly mentioning recent developments and directing to recent reviews on the topic a suitable compromise.

10. *The limitations section would benefit from further elaboration, especially regarding the challenges of studying intergenerational health transmission over such an extended time frame. Consider discussing the potential for cohort effects or other time-sensitive variables (e.g., advancements in healthcare or neonatal care) that could influence the outcomes across generations.*

Thank you for raising this concern. Although we provided 'overlapping generation' analyses comparing survival in G3 and G4 descendants to assess the role of such cohort effects and other time-sensitive variables (Lines 177-183, 205-214, 279-292, 397-402, 420-434), we now highlight that our approach may be susceptible to residual biases (Lines 526-528).

11. *It would also be useful to briefly outline future research directions that could address these limitations, such as prospective studies of health outcomes in G4 great-grandchildren as they age.*

Thank you for raising this excellent point. In the discussion section of the manuscript, we provide an explicit discussion of these research avenues (Lines 532-535).

Reviewer 2

Reviewer #2 (Remarks to the Author):

I have once again revised this manuscript. As I detailed in my earlier, more extensive critique of the first submission—and reiterated in the last one—I stated that even in explanatory research, cross-validation is required to assess model stability, generalizability, and overfitting, particularly in the context of large sample sizes. In this review, I only address my previous concerns and have not assessed the overall manuscript in other aspects.

Once again, the authors have refused to perform these analyses, arguing that "these methods are not able to produce accurate statistical inference (p-values, confidence intervals) on their estimated covariate parameters," which is simply not true. Their reluctance to conduct the requested analyses, combined with the flawed argument they have repeated, raises doubts about the quality of the work.

The authors argue that penalized regression models (e.g., LASSO) are not suitable for their explanatory research aims. However, I disagree entirely. my concern remains that cross-validation can help verify model stability even in explanatory models. The response includes a sensitivity analysis using penalized Cox models with cross-validation (via the glmnet package in R). Still, it is unclear why they did not perform the requested analysis. This omission further increases doubts about the robustness of their results.

Other Points

Explicit mention of limitations in the abstract:

The response states that the wording was adjusted in the abstract to indicate limitations in measuring socioeconomic factors. However, I explicitly suggested adding a clear limitations statement, which has still not been done.

Over-reliance on non-significant trends: I was concerned about interpreting results that were not statistically significant but were used as evidence of trends. The response explains that composite outcomes were used to strengthen conclusions. However, they should report only global scores in such cases; otherwise, they must explicitly state that non-significant effects cannot be used as evidence. Such patterns should not be over-interpreted in the absence of statistical significance.

Dear Reviewer 2,

Thank you for your review of our study. We are sorry that you deem our previous responses inaccurate and incomplete. In the present response, we endeavored to clarify some points that may have been missed or miscommunicated previously and provide additional analyses to alleviate methodological concerns. During the writing of the present response, we consulted the expertise of our colleague Sören Möller, Professor of Biostatistics. We thank him for his excellent discussion, and he is happy to be acknowledged for his contribution both here and in our manuscript.

1. *Explicit mention of limitations in the abstract: The response states that the wording was adjusted in the abstract to indicate limitations in measuring socioeconomic factors. However, I explicitly suggested adding a clear limitations statement, which has still not been done.*

Thanks for you for this comment. We agree, and we have added a mention of the limitations of the socioeconomic measures we have employed in our study, which has increased the word count beyond 150 words to 163 words. We hope that you and our editor find this change suitable. New abstract with the limitations statement highlighted:

> "Descendants of longevity-enriched sibships demonstrate a broad health and survival advantage throughout the life course. However, little is known about manifestations during very early life. Here we show a pattern of lower risk of adverse early life outcomes in third-generation grandchildren (N = 5637) of Danish longevity-enriched sibships compared to the general population, including infant mortality (Hazard Ratio = 0.53, 95% CI [0.36, 0.77]) and a range of neonatal health indicators. These associations in fourth-generation great-grandchildren (N = 14,908) were attenuated and less consistent (e.g., infant mortality, Hazard Ratio = 0.90, [0.70, 1.17]). These dilatory patterns across successive generations were independent of stable advantages in select socioeconomic and behavioral indicators (e.g., parental education, income and maternal smoking), maternal and paternal lines of transmission, as well as secular trends in the background population. However, our socioeconomic and behavioral indicators were limited in both range and granularity. Our findings suggest that the familial aggregation of exceptional health and survival has early life developmental components."

2. *Over-reliance on non-significant trends: I was concerned about interpreting results that were not statistically significant but were used as evidence of trends. The response explains that composite outcomes were used to strengthen conclusions. However, they should report only global scores in such cases; otherwise, they must explicitly state that non-significant effects cannot be used as evidence. Such patterns should not be over-interpreted in the absence of statistical significance.*

Thank you for this comment. We agree, and we have removed the following sentence from the first paragraph of the discussion section, to avoid overinterpreting non-significant trends:

"*The infant health advantage observed in our study manifested with protective trends across a broad range of outcomes, suggesting that survival differences were not driven by protection against a narrow range of conditions related to a particular aetiology*."

3. *Once again, the authors have refused to perform these analyses, arguing that "these methods are not able to produce accurate statistical inference (p-values, confidence intervals) on their estimated covariate parameters," which is simply not true. Their reluctance to conduct the requested analyses, combined with the flawed argument they have repeated, raises doubts about the quality of the work.*

We would like to apologize for our previous response on this matter. We began by reiterating our initial concerns with respect to the recommended methodological approach, which may have given the impression that we still had not performed the requested analyses. However, we would like to clarify that we did perform analyses that addressed the issues raised in previous reviews, namely: use of robust penalized regression analyses, K-fold cross validation, and explicitly assessing whether our main models without penalization were subject to overfitting or lack of prediction generalization. We did this by employing the '***glmnet***' package in R, which has implemented Elastic Net, Ridge or Lasso regularization for a range of generalized linear models.

However, we did not state in our previous response what type of penalized regression method we used, which may have also been a point of confusion. We employed Elastic Net regularization, whilst reviewer 2 had previously suggested Ridge or Lasso. We apologize for this oversight, and to remedy this, we have updated our general population analyses described previously to include, in addition to Elastic Net regularization, results for Lasso and Ridge models as well. As in our Elastic Net models, these new models show the same patterns of there being negligible risk for overfitting in unpenalized compared to penalized models. Specifically, we observe that:

A. The estimates for infant mortality were similar in both the unpenalized methods and the penalized methods. This suggests that the parameters were not penalized due to overfitting.
B. The optimized regularization parameter (λ) determined by 10-fold cross-validation in penalized models was very close to zero (λ = 0 is equivalent to non-penalized regression), implying that there was negligible risk of overfitting in the unpenalized models.
C. The cross-validated Harrel's C index, our main performance prediction metric, was consistent across unpenalized and penalized models. This suggests that the out-of-sample performance of non-penalized models was equivalent to that of the penalized models, supporting prediction generalizability even without cross-validation to optimize model fit, suggesting a negligible risk of overfitting.

We have summarized the results of these analyses, as well as our previous analyses, in a series of tables in the Supplementary Appendix, under item S3.14.

4. *(The authors) … argue that "these methods are not able to produce accurate statistical inference (p-values, confidence intervals) on their estimated covariate parameters," which is simply not true.*

We kindly disagree with reviewer 2 about the accuracy of this statement – our original quote was that *"**most models of these varieties** are not able to produce accurate statistical inference (p-values, confidence intervals) on their estimated covariate parameters"*. This constitutes an active area of research, and there exists little consensus on how to tackle this issue in practice. Most implementations of penalized regression models (e.g. Ridge, Lasso, Elastic Net regularization) in various statistical software often deliberately avoid reporting statistical inference on parameter estimates because of this, including '***glmnet***'. Chapter 6 of the vignette for the '***Penalized***' package in R has an explicit explanation for why in their case: https://cran.r-project.org/web/packages/penalized/vignettes/penalized.pdf). The difficulty of producing statistical inference for such models arises for the following reasons:

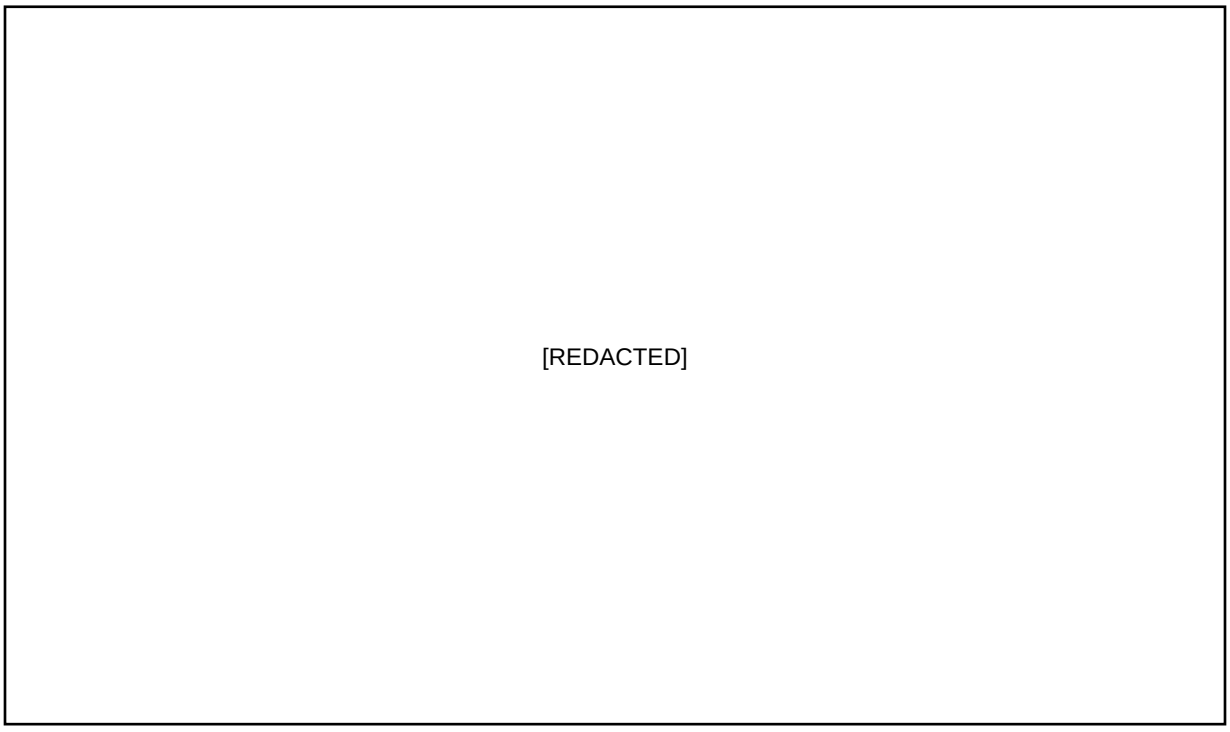
A. The partial likelihood used in penalized regression models cannot be differentiated and therefore p-values and confidence intervals, in general, cannot be calculated parametrically from known distributional properties of these models.
B. Although bootstrapping methods for penalized regression methods are technically possible, they are generally not meaningful. Broadly penalized parameter estimators (e.g. Ridge or Elastic Net) are generally no longer unbiased with respect to their assumed population estimands (relative to the severity of the penalization), and so bootstrapped p-values and confidence intervals would not have the expected frequentist interpretations regarding the relationship between the estimate and the population estimand and may be misleading.
C. It is technically possible to use standard statistical inference for unpenalized methods subsequent to variable selection with Lasso regression (which removes variables from the model rather than broadly penalizing parameters). However, without special consideration, this approach is generally not valid due to the fact that data is used twice in the estimation procedure (see below, re selectiveInference).

There exist a few R implementations of methodologies that attempt to circumvent these issues, although none do so without limitations, and are viewed as preliminary developments. The package '***glmnetSE**'* by an author unaffiliated with '***glmnet**'* implements bootstrapped standard errors on penalized models derived from the '***glmnet***' package. However, it can only provide standard errors for estimates on parameters that are specifically exempt from the penalization set, and the interpretative issues described previously remain. Moreover, this can only be done for GLM models in the gaussian and binomial families and is currently unavailable for Cox models. The package '***selectiveInference***' by Tibshirani *et al.* employs modified bootstrapped confidence intervals in the context of variable selection with Lasso regression. The main limitation of this approach is that inference is conditional on the penalized model and so has a different interpretation, and is only valid for *a priori* chosen lambda values and not those determined by cross-validation. In the context of Lasso Cox proportional hazard regression models where lambda is determined by cross-validation, the authors state: "*Type I error control is good, except in the cross-validation case where it is badly anti-conservative for smaller P-values*" (DOI: 10.1002/cjs.11313).

Since glmnetSE cannot be used for Cox proportional hazard regression models, we could not implement this on our data. However, we have implemented the approach from the '***selectiveInference'*** package using the optimal lambda value determined from our previous Lasso regression analyses. As described above, using cross-validation to determine the optimal lambda affects the accuracy of this approach, which is one of its main limitations. These results can also be found under item S3.14.

5. *As I detailed in my earlier, more extensive critique of the first submission—and reiterated in the last one—I stated that even in explanatory research, cross-validation is required to assess model stability, generalizability, and overfitting, particularly in the context of large sample sizes.*

We would like to bring attention to the repeated assertion that cross-validation is particularly needed in the context of large sample sizes. If reviewer 2 is implying overfitting becomes worse in large sample sizes, then we kindly disagree with this statement. The opposite is true - for a fixed number of covariates, out-of-sample prediction accuracy always improves with larger sample sizes. The most extreme case of overfitting occurs when the number of parameters is equal to the sample size and the model perfectly interpolates the data points. This situation occurs either in extremely high-dimensional settings or in low sample sizes. This point and the relationship between sample size and risk of overfitting is nicely illustrated in Figures 1 and 2 of a *Nature Methods* article on overfitting (DOI: 10.1038/nmeth.3968). These figures are copied below for ease of viewing.



[REDACTED]