Reviewer 1

1. *The authors combined 6 survey cycles (12 years). As stated in lines 238 -241, and based on the provided R code, they didn't use the appropriate weight. As stated by NCHS guidelines, the authors should have created a combined sample weights "MEC12YR" of the 6 survey cycles (12 years), which would be MEC12YR= 1/6 *WTMEC2YR (for cycles 1999-2000, 2001-2002, 2003-2004, 2005-2006, 2007-2008, 2009-2010). The authors should refer to the NHANES Tutorial https://wwwn.cdc.gov/nchs/nhanes/tutorials/module3.aspx, and re-run the analyses using the appropriate combined weight.*

**RESPONSE: We thank the Reviewer for pointing out our previous error in the sample weights used. Per the NHANES tutorial linked to by the Reviewer, MEC12YR should be calculated as (2/6) * WTMEC4YR for cycles 1999-2000 and 2001-2002, and as (1/6) * WTMEC2YR for all subsequent cycles. All tables and figures have been updated to reflect analyses with these updated sample weights. Importantly, neither the statistical significance nor the findings of the manuscript are altered by the use of the combined sample weights.

We have added the following to the Methods: "We followed all NHCS guidelines for the analysis of NHANES data 62. As the survey weights relevant to the smallest sample subpopulation for which all data are available should be used, we used mobile examination center (MEC) weights to adjust for complex survey design, oversampling, non-coverage, day of the week, and survey nonresponse to compute nationally representative estimates 63,64. Per NHANES analytical guidelines for combining data across cycles, 12-year MEC weights were calculated using the NHANES-provided variables WTMEC4YR and WTMEC2YR as follows: WTMEC12YR=1/3*WTMEC4YR for the 1999-2000 and 2001-2002 cycles and WTMEC12YR=1/6*WTMEC2YR for all subsequent cycles."

Reviewer 2

1. *Provide more specific information on relationships of parity to insulin resistance models in C. elegans and mitochondrial theory of aging (beyond just free radicals).*

**RESPONSE: We thank the Reviewer for this comment. We acknowledge that the previous version of our manuscript did not reference specific proximate mechanisms beyond epigenetic changes and telomere shortening that have been proposed to accelerate biological aging. We have added the following to the first paragraph of the paper: "This variation - attributed to differences in the biological rate of deterioration or repair - is referred to as 'biological age', and is thought to reflect the cumulative effect of environmental exposures in combination with underlying genetic variation. Various proximate mechanisms have been proposed to modulate biological age acceleration, including insulin signaling [4], oxidative stress [5], inflammation [6], epigenetic changes [7], and telomere shortening [8]. Understanding the environmental, behavioral, and physiological factors that influence biological aging may inform policies and interventions that could help to mitigate their effects, thereby extending the healthspan."

2. *Number the Mahalanobis (1936) and Levine (2013) references, plus the Liu et al. reference on lines 205-206.*

**RESPONSE: References have been added as numbers in the mentioned instances.

3. *List the demographic and other NHANES variables used. For example, the authors describe computation of BMI, but it already exists as a variable in NHANES: BMXBMI.*

RESPONSE: We thank the Reviewer for this comment. We have updated the Method in two sections to specify which NHANES-provided variables were used for analysis, described below.

The section *Reproductive health and parity data* now reads as follows: "Women completed a computer-assisted questionnaire on their reproductive health history. Women reported whether they were currently pregnant, if they have ever been pregnant, how many pregnancies resulted in a live birth (if applicable; NHANES items RHD170 and RHQ171), whether they had regular periods over the last 12 months, and their reason for not having regular periods over the last 12 months (if applicable). As previous work has suggested that current pregnancy modulates certain measures of biological age [14], women who self-reported currently being pregnant were excluded from analyses (NHANES item RIDEXPRG; n = 1,417 out of all women between 18 and 84)."

In the *Covariates* section, we have now added the following: "Self-reported race/ethnicity [58], socioeconomic status (SES) [59,60], and smoking [10] moderate the relationship between chronological age and biological aging. Self-reported race/ethnicity was categorized as non-Hispanic (NH) white, NH black, Hispanic, and 'other' (NHANES item RIDRETH1). SES was indexed by educational attainment (NHANES item DMDEDUC2) and federal income-to-poverty ratio (FIPR; NHANES item INDFMPIR as calculated per Department of Health and Human Services guidelines). Height and weight were measured by an NHANES examiner, and BMI was calculated as weight (kg) divided by height (meters squared; NHANES item BMXBMI)."

4. *Given multiple variable comparisons, adjust down the Type I Error rate using Bonferroni's method.*

RESPONSE: We thank the Reviewer for this comment. We have added the following to the Statistical analysis section: "Because we estimated four regressions (one per outcome measure) for each set of analyses for each analytical subset, statistical significance was set to $p < 0.0125$ $(0.05/4)$ [65]."

In the results section, we now specify that any results discussed as being statistically significant were significant after correcting for multiple comparisons.

5. *Besides the Figure 3 description for the linear regression estimates, spell out the equations for each linear regression model that was tested.*

RESPONSE: We have now added ESM Text I, which includes the regression equations for our primary analyses and sensitivity analyses. ESM Text I reads as follows:

ESM Text I. Regression equations for primary and sensitivity analyses

Primary model
Predicted biological aging measure = $b_0 + b_1$(live births) + b_2(live births^2) + b_3(age) + b_4(BMI) + b_5(BMI^2) + b_6(FIPR) + b_7(smoking) + b_8(education) + b_9(ethnicity) + error
Sensitivity analysis 1
Predicted biological aging measure = $b_0 + b_1$(live births) + b_2(live births^2) + b_3(age) + error
Sensitivity analysis 2
Predicted biological aging measure = $b_0 + b_1$(live births) + b_2(live births^2) + b_3(BMI) + b_4(BMI^2) + b_5(FIPR) + b_6(smoking) + b_7(education) + b_8(ethnicity) + 9(years since last live birth) + b_{10}(years since last live birth)(live births) + b_{11}(years since last live birth)(live births^2) + error

Sensitivity analysis 3
Predicted biological aging measure = b_0 + b_1(live births) + b_2(live births^2) + b_3(months since last live birth) + b_4(age) + b_5(BMI) + b_6(BMI^2) + b_7(FIPR) + b_8(smoking) + b_9(education) + b_{10}(ethnicity) + b_{11}(live births)(months since last live birth) + b_{12}(live births^2)(months since last live birth) + error

6. *Limitations are discussed, but the various measures (e.g., cytokines, immune cells) represent snapshots in time that might be explained by other unmeasured conditions (e.g., mild infections) that participants happened to be experiencing at the time of their NHANES physical and laboratory assessments. Address this limitation. This goes beyond what the authors describe in terms of pregnancy effects that may or may not be captured during the participants' NHANES assessments.*

**RESPONSE: We thank the reviewer for this point, and agree that NHANES data represents a cross-sectional view of someone's condition which may certainly be affected by unmeasured conditions present at the time of assessment. Acknowledging this point, we have added the following to the *Limitations* section: "In the absence of longitudinal sampling, we also cannot be certain that biomarkers measured in this cross-sectional sample are not also representative of transient states unrelated to parity or reproduction. For example, it is possible that some participants could have been experiencing mild infections during MEC examinations, leading to altered clinical measures of immune function. Though this could contribute to imprecision in our biological aging measures, such imprecision would not be systematic and thus we would not expect it to significantly affect the present study's findings."

7. *Did the authors also consider using various measures of allostatic load or American Heart Association Cardiovascular risk using NHANES variables as proxies for aging?*

**RESPONSE: We appreciate the Reviewer's suggestion to include allostatic load in our manuscript. In light of this suggestion, we have added an implementation of allostatic load since it also approximates cumulative physiological dysregulation across multiple organ systems. This iteration was constructed using the same panel of biomarkers as the metrics included in the original submission and use risk thresholds defined by quartiles within the sample per recommendations of Duong et al., 2017, *Journal of Racial and Ethnic Health Disparities, 4:*455-461, in their review of different allostatic load implementations in NHANES.

We have added the following to the introduction: "Four composites of system integrity have been used to operationalize biological age and cumulative system dysregulation within the context of large-scale epidemiological studies in the United States: Homeostatic Dysregulation (HD) [45], Levine Method Biological Age (LM) [46,47], the Klemera-Doubal Method Biological Age (KDM) [46,48], and allostatic load (AL) [49]. Previous work using a nationally representative sample of adults in the US from the National Health and Nutrition Examination Survey has found that HD, LM, KDM, and AL exhibit robust associations with physical functioning, cognition, hearing and vision, and with self-reports of health and functional disability [11,50]. Other population-based studies have found similar links between AL and both objective and subjective markers of physical functioning and general health[51]."

We have added the following to the section Biological aging measures: "Allostatic Load (AL) is computed as the proportion of biomarker values for which a participant is at risk. In accordance with recommendations from a review of AL implementation in NHANES [54], we defined risk as residing within the highest quartile of a given biomarker's distribution within the sample of nonpregnant women aged 18-84 with complete biological age biomarker data, excepting

albumin for which risk was defined as residing in the lowest quartile (N = 5,870; **ESM Table VI**). In this manner, the number of biomarkers for which a participant is at risk is divided by the total number of biomarkers in the panel to calculate a final allostatic load score with values ranging from 0-1."

Regressions with allostatic load as an outcome variable have been added. As seen in Table 2, allostatic load exhibits similar relationships with live births as do the other biological aging measures.

8. *Provide stronger evidence for validity of these composite measures and how they are appropriate in NHANES.*

RESPONSE: We have clarified that these composite measures have been validated *specifically* using NHANES data previously. We have added the following to the introduction: "Four composites of system integrity have been used to operationalize biological age and cumulative system dysregulation within the context of large-scale epidemiological studies in the United States: Homeostatic Dysregulation (HD) [45], Levine Method Biological Age (LM) [46,47], the Klemera-Doubal Method Biological Age (KDM) [46,48], and allostatic load (AL) [49]. Previous work using a nationally representative sample of adults in the US from the National Health and Nutrition Examination Survey has found that HD, LM, KDM, and AL exhibit robust associations with physical functioning, cognition, hearing and vision, and with self-reports of health and functional disability [11,50]. Other population-based studies have found similar links between AL and both objective and subjective markers of physical functioning and general health[51]."

9. *Given the need for longitudinal studies that the authors cite, have they considered applying these models to longitudinal data sources such as the Baltimore Longitudinal Study on Aging or the MacArthur Studies of Successful Aging?*

RESPONSE: We thank the Reviewer for this point and certainly agree with both the need for longitudinal studies, and with those studies as potential fruitful avenues for future work. For the present study, we opted to use the cross-sectional NHANES dataset because the biological age measures had been previously validated to track closely with both subjective and objective measures of health and physical functioning, increasing our confidence in the validity of these biological age measures (see response to point 8 above).

Reviewer 4

1. *My big issue is the inclusion of adjustment for chronological age in the multivariable models. I appreciate the utility of assessing the correlation of the measures of biological age with chronological age to substantiate that they are measuring something of the same thing. However, they are highly collinear, albeit to a lesser extent HD, and two of the measures – KDM and LM – incorporate chronological age in the construct, so it seems inappropriate to adjust for this in your main models.*

RESPONSE: We agree with the reviewer's concern about the collinearity between chronological age and biological age in multivariate models. To address this concern, we conducted all analyses using versions of each measure that had been adjusted for chronological age, but did not include a description of these steps in the original methods. Our approach described in the text below has been added to the *Statistical Analyses* section of the Methods: "To focus on biological aging, we conducted analyses using versions of each biological age measure after adjustment for chronological age, computed as the residuals of

We have also added the following to the same section:

2. *Another issue I have is the model itself and its use of quadratic and interaction terms. The authors argue the inclusion of quadratic terms for parity and BMI is because they have been suggested to have a U-shape in their associations with health outcomes. While this may be so, there are better ways to assess this than adding a quadratic term. My preference would be to use a categorical term. For parity this is easy, it's just a 0-6 term. For BMI, you could use the WHO categories (underweight/normal/overweight/obese). In this fashion you can evaluate shape while allowing an easier interpretation of the model. I know statisticians are disinclined to categorise because you lose information, but I think these suggested terms are a good compromise.*

RESPONSE: To quantitatively assess whether data follow a U-shape, it is necessary to include both linear and quadratic terms in a model and as such, we chose to include linear and quadratic terms for both BMI and number of live birth (see also Montgomery et al. 2012, "Introduction to Linear Regression Analysis"). As Reviewer 4 points out, transforming continuous data to categorical data loses information, and also reduces model degrees of freedom. As such, we have chosen to present models with BMI represented as a continuous variable rather than a categorical variable.

To address Reviewer 4's concern about BMI, we include a variable "BMI_cat" in our data file, which follows the recommendation of the reviewer to group BMI into categories per the WHO guidelines, and sets the 'normal' category as the reference category. We also include in our code file lines to run our primary regressions when substituting continuous BMI with categorical BMI. Importantly, the statistical significance and interpretation of our results with respect to parity and menopause (and their interactions) are not affected by the decision to treat BMI as a continuous or categorical variable.

Categorizing number of live births as an ordinal rather than continuous variable would similarly require choosing a parity as the reference category to which all other number of live births would be compared to. The resulting model would provide estimates of the difference between the reference category and all other categories of live births. As we have 8 number of live birth categories (0 to 7 live births, inclusive) this would mean one category would have to arbitrarily be set as the reference category, and all estimates (of which there would be 7) would only describe how each other category compares to the reference category. A categorical model like the one described would not allow us to statistically test whether the data conform to a U or J-shaped distribution. In our view, this approach would be inappropriate and lead to a significant loss of information relative to the employed approach of treating the number of live births as a continuous variable.

3. *For the interaction terms, the authors include these for live birth number vs menopause, again because prior literature suggests a difference by menopause state. While this may be so, I'm disinclined to just include an interaction term in the model because 1) you're not really showing the interaction and 2) you can't then interpret the coefficients directly because they're*

partly expressed in the main term and in the interaction term. Instead, if you want to assess interaction present stratum-specific estimates of your main association with a test of difference.

RESPONSE: We have updated our analyses to be stratified by menopausal status, as first described in the Statistical Analyses section: "As prior work suggests that costs of reproduction should be the most apparent after menopause [44,66], models were estimated separately premenopausal and post-menopausal women."

Our results are now broken up into separate sections for pre- and postmenopausal women and read as follows: *Differences between pre-menopausal and post-menopausal women* Demographic differences and differences in biological age acceleration are presented in Table 1. When adjusting for demographic differences, pre-menopausal women exhibited significantly lower LM and KDM biological age acceleration relative to post-menopausal women.

Pre-menopausal women
The linear effect of number of live births and squared term, or quadratic effect, of live births was not significant in any primary model in pre-menopausal women ($n = 2,166$; see **Table 2; Figure 4**). Sample sizes for our sensitivity analyses controlling for chronological age only were slightly larger ($n = 2,686$), as less participants were excluded due to missing covariate information. Similar to our primary analyses, the main effects of live births (both linear and quadratic terms) were not significant across all measures of biological age (**Table 2**). Repetition of these analyses in the primary analytical sample yielded the same pattern of results. Of the 2,166 premenopausal women in our primary analyses, data on years since last live birth were available for 1,617. The average years since last live birth was 8.87 (SE = 0.19). After correcting for multiple comparisons, the main effect of years since last live birth was not significant in any model, nor were any of the interaction terms between years since last live birth and parity (**Table 2**).

Our sample size for analyses including months since last live birth ($n = 107$) was significantly limited by the fact that this subsample excluded all post-menopausal women, and excluded women sampled prior to this question being added in the 2007-2008 cycle. On average, women with valid responses to this question gave birth 10.7 months ago (SE = 0.63). After correcting for multiple comparisons, the main effects of months since last live birth and parity was not significant in any model, nor were any of the interaction terms between months since last live birth and parity (**Table 2**). These results should be interpreted with caution given the small sample size.

Post-menopausal women
Primary models in post-menopausal women revealed a significant linear effect of live births on biological aging indexed by LM, HD, and AL; the linear effect of live births on KDM was not significant after correction for multiple comparisons ($n = 2,252$; **Table 3**). After correcting for multiple comparisons, the quadratic effect of parity on biological aging was significant for all measures but KDM. Sample sizes for our sensitivity analyses controlling for chronological age only were slightly larger ($n = 2,498$). Similar trends were observed in the first set of sensitivity analyses, wherein the linear effect of live births was significantly associated with LM, HD, and AL. Moreover, the quadratic effect was significant for all four measures, giving rise to the anticipated U-shape for the overall relationship between parity and biological aging (shown in grey on **Figure 4**). Repetition of these analyses in the primary analytical sample yielded the same pattern of results. Of the 2,252 post-menopausal women in our primary analyses, data on years since last birth were available for 1,970. The average years since last birth was 36.09 (SE = 0.25). After correcting for multiple comparisons, the main effect of years since last live birth

was not significant in any model, nor were any of the interaction terms between years since last live birth and parity (**Table 3**).

4. *The Introduction is a bit confusing with regard to the three measures of biological age here and the telomeric and methylation-based measures. In the first mention it insinuates the telomere/methylation are better but too costly. Then later it's stated that all the measures are assessing different aspects and don't really correlate. This is a bit peculiar because one presumes that telomere/methylation correlate with age and all the measures of biological age here correlate with age (HD not so much), so how they are not then correlated with one another is puzzling. In any event, some reconciliation of whether the telomere/methylation measures of age are worthwhile would be good.*

**RESPONSE: We thank Reviewer 4 for encouraging us to expand on the different measures of biological aging and the relationships between them. We have edited the Introduction to clarify why physiological-based measures of biological aging might better reflect costs of reproduction as follows: "Cellular measures of biological age such as telomere length and DNA methylation age may provide insights into the molecular processes linking reproduction to mortality and other health outcomes [7,38], and may eventually serve as early indicators of the costs of reproduction in health and aging. However, 'aging' may refer to a wide range of processes that may occur at different times or at different speeds. For example, cellular measures of biological age that examine mitotic (e.g., telomere length) and non-mitotic (e.g., DNA methylation age) processes are often weakly correlated [14,39]. Similarly, both telomere length and DNA methylation age are poorly associated with measures of biological age implemented at the clinical level [40–43]. Thus, it has been suggested that different measures of biological age and cumulative system dysregulation index fundamentally different components of the aging process.

Clinical measures of biological age quantify changes in physiological integrity by combining information from multiple clinical biomarkers that collectively assess the functioning of major organ systems throughout the body. Such measures may be particularly relevant in light of the many physiological, immunological, and endocrinological changes that accompany reproduction in women [44]. Four composites of system integrity have been used to operationalize biological age and cumulative system dysregulation within the context of large-scale epidemiological studies in the United States: Homeostatic Dysregulation (HD) [45], Levine Method Biological Age (LM) [46,47], the Klemera-Doubal Method Biological Age (KDM) [46,48], and allostatic load (AL) [49]. Previous work using a nationally representative sample of adults in the US from the National Health and Nutrition Examination Survey has found that HD, LM, KDM, and AL exhibit robust associations with physical functioning, cognition, hearing and vision, and with self-reports of health and functional disability [11,50]. Other population-based studies have found similar links between AL and both objective and subjective markers of physical functioning and general health[51]. Clinically-based measures may therefore provide an affordable and accessible alternative to cell-based measures for measuring systemic deterioration tied to costs of reproduction in women."

5. *The three measures of biological age use different groups of women in the reference population, both in number and age range. HD uses 482 women aged 20-30 with non-obese BMI, KDM uses women aged 30-75, and LM seems to just use them all, though I'm not sure. It seems in having more women to draw from would give a superior estimate. Also, unclear about the inter-comparability of the three biological measures if they're drawing from such different populations. The authors have greater familiarity with these measures than I so I presume it's ok but perhaps some text in the paper to assuage such concerns would be good.*

RESPONSE: The reviewer's intuition is correct that differences in the reference population and approach toward scale construction reflect differences in theoretical conceptualizations of biological aging. To assuage concerns that these decisions are not made haphazardly we have added the text below to the methods in an effort to relay how the various approaches employed reflect these conceptual and theoretical differences: "All four biological aging measures were computed using the same panel of 9 biomarkers. These biomarkers were selected based upon their inclusion in the LM biological age algorithm, which utilized machine-learning analysis to select the most parsimonious panel of biomarkers for mortality prediction. The use of common biomarkers ensures the different measures are indexing the same physiological processes. Differences in the analytical approach and statistical operations leading to the final composite measure reflects different approaches toward the conceptualization of biological age. For HD, biological age is conceptualized as deviation from an ideal physiological state attained in one's 20s. For KDM, biological age is conceptualized as the average change in physiology that occurs with increasing chronological age. Building upon this, LM captures the increased risk in mortality that accompanies physiological changes occurring with age. Finally, AL conceptualizes aging as the accumulation of changes that become impactful only once they reach a critical threshold. Biomarker and biological age summary statistics for the final analytical sample ($n = 4,418$) are provided in **ESM Table VIII.**"

6. *One thought I had was the restriction to live births. If the hypothesis is that live births put such a strain on the body that it is deleterious to lifespan then shouldn't all pregnancies have this effect? Admittedly there's a complexity as to when the pregnancy terminated, either by termination or miscarriage, but this could be an acknowledge limitation. Alternatively, you could include all pregnancies and then do a further analysis restricted to live births which presumably would cluster around 9 months.*

RESPONSE: We thank the Reviewer for this point. The energetic and physiologic cost of pregnancy varies significantly across gestation, with energetic costs during early gestation being minimal as compared to during the third trimester. Because NHANES does not collect information on the duration of each reported pregnancy, and because of the prevalence at which clinically recognized pregnancies result in natural miscarriages, we chose to analyze live births rather than pregnancies. We have now acknowledged this in the *Method* as follows: "NHANES does not collect fine-grained data about pregnancies that do not result in live births, rendering it impossible to estimate the length of each pregnancy, and concomitantly, the physiological cost of each pregnancy. Further, approximately 30% of implantations end in natural miscarriage [53], making number of recognized pregnancies a more imprecise measure of physiological investment in reproduction as compared to number of live births. As a result, we chose to use number of live births rather than number of pregnancies."

We have also added the following to the Limitations section: "Another limitation is our reliance on the relatively crude measures of reproductive effort in women. We were restricted to a measure of life births, but do not have access to data on miscarriages or aborted pregnancies, which could also be associated with costs of reproduction. We also lack information on breastfeeding, which is energetically costly in women [79]. Nevertheless, the fact that we do detect a strong and robust signal of accelerated biological aging with parity in post-menopausal women implies that parity is adequate to capture important health-related costs in this population."

7. *Authors suggest impacts of pregnancy may be chronic and so they examine the duration since last birth. If it's a chronic effect, however, shouldn't it be duration since first birth?*

RESPONSE: We thank the Reviewer for an opportunity to clarify these analyses, as the manuscript's introduction previously did not describe them clearly. The set of analyses including duration since last birth aim to "assess the extent to which effects of parity may be chronic and accumulate over time, or acute and only present in the postnatal period" (see *Sensitivity Analyses*). We have added the following to the introduction: "Based on findings from the most highly powered prior studies of all-cause mortality and parity, we hypothesized a U-shaped relationship between parity and biological aging. Specifically, we predicted that accelerated biological aging would be most apparent in women with the lowest and the highest parity. We also leverage this powerful dataset for preliminary tests of whether relationships between parity and biological age are durable, such that they persist regardless of time since last birth, or transient, such that the effect of parity on biological age decreases as a function of time since last birth."

We have also added the following to the Discussion: "That parity was not associated with biological age in pre-menopausal women along with the fact that time since last birth did not predict biological age acceleration in either pre- or post-menopausal women supports the argument that the effects of parity are durable, and not simply short-term physiological changes associated with pregnancy and breastfeeding."

8. *The sensitivity analyses are stated to be in a larger sample (n=3,235). To aid in comparability with the main analyses, should present results constrained to the same sample. If no difference, you can just state that.*

RESPONSE: We have added the following to the Results section on premenopausal women: "Similar to our primary analyses, the main effects of live births (both linear and quadratic terms) were not significant across all measures of biological age (**Table 2**). Repetition of these analyses in the primary analytical sample yielded the same pattern of results."

We have added the following to the section on postmenopausal women: "Similar trends were observed in the first set of sensitivity analyses, wherein the linear effect of live births was significantly associated with LM, HD, and AL. Moreover, the quadratic effect was significant for all four measures, giving rise to the anticipated U-shape for the overall relationship between parity and biological aging (shown in grey on **Figure 4**). Repetition of these analyses in the primary analytical sample yielded the same pattern of results."

9. *The flowchart in Figure 1 shows marked attrition on restriction to the sample with complete biomarkers, going from 92,062 to 8,130. How representative are these 8,130 of the parent cohort?*

RESPONSE: The reviewer is correct that our data restriction from the full sample to those with complete biomarker data is quite significant. However, the numbers originally displayed in Figure 1 are an exaggeration of this process. First, the 92,062 reflects both male and female participants and is inclusive all the way to NHANES 2015 and the original 8,130 participants includes those under age 18, which are not included in our analytical sample. Thus, we have edited Figure 1 to better reflect the decision points in our sampling reduction by showing how the sample was reduced from all female participants in NHANES 1999 - 2010 (N=31,575) to all non-pregnant women aged 18-84 (N=13,929). We consider this subset of 13,929 women to comprise the largest hypothetical sample, since restriction from this point onward was dictated by data availability (as described in text and Figure 1). Thus, to determine representativeness we compared characteristics of this sample to the subset of non-pregnant women aged 18-84 with complete biological age biomarker data (N=5,870). The results of these analyses are

presented in **ESM Table I**. Here we noted no significant differences in age, ethnicity, education, income, smoking status, or menopausal status between the sample of all non-pregnant women aged 18-84 and those with complete biomarker data. However, the sample with complete biomarker data was significantly more likely to have been pregnant before, a consolation we are satisfied with given the aims of the work. We have added a brief review of these comparisons in the "Data source" section of the Methods: "To assess the representativeness of participants with complete biomarker information, we compared the subset of non-pregnant women aged 18-84 with complete biomarker data (n = 5,870) to all non-pregnant women aged 18-84 in NHANES 1999-2010 (n = 13,929). The two samples were similar in age, ethnicity, educational attainment, income, smoking status, menopausal status, and number of live births. However, the sample with complete biomarker data was significantly more likely to have ever been pregnant. Comparative demographics and associated tests of difference are reported in **ESM Table I**."

10. *In the Methods, please specify how many pregnant women were excluded.*

RESPONSE: We have added the following to the Method: "As previous work has suggested that current pregnancy modulates certain measures of biological age [9], women who self-reported currently being pregnant were excluded from analyses (NHANES item RIDEXPRG; n = 1,417 out of all women between 18 and 84)."

11. *If the parity number is capped at 7, I think the text about top-coding at 11 is unnecessary.*

RESPONSE: We have removed this from the manuscript.

12. *Please add a reference for the FIPR.*

RESPONSE: We have added the following to our initial discussion of FIPR in the Method: "Self-reported race/ethnicity was categorized as non-Hispanic (NH) white, NH black, Hispanic, and 'other' (NHANES item RIDRETH1). SES was indexed by educational attainment (NHANES item DMDEDUC2) and federal income-to-poverty ratio (FIPR; NHANES item INDFMPIR as calculated per Department of Health and Human Services guidelines)."

13. *Legend for Figure 4 reports a different model from that in Table 2.*

RESPONSE: Figure 4 now displays the results of models in Table 2. The caption for Figure 2 now reads as follows: "Predicted values and 95% confidence intervals derived from primary models for LM age acceleration (panel A), HD acceleration (panel B), KDM age acceleration (panel C), and AL age acceleration (panel D) among premenopausal women (black lines) and postmenopausal women (grey lines), National Health and Nutrition Examination Survey (n = 4,418)."

14. *In presenting data in Table 1, please present continuous as Mean (SD; range) and di/polytomous as n (%).*

RESPONSE: Data for continuous variables are now reported as Mean (SE, range) and categorical variables are reported as n (%).

15. *If you could, would be helpful to see the summary statistics for the three biological age outcomes.*

RESPONSE: These statistics have been added alongside the biomarker summary statistics for the full sample in **ESM Table VIII**. Summary statistics for biological aging measures as a function of menopausal status are presented in **Table 1**.

16. *Note in Table 2 that Other Race/Ethnicity is not there. Were they dropped or aggregated with another group?*

RESPONSE: Other race/ethnicity is now displayed in Table 1.

17. *For Table 2, would prefer having beta coefficient and 95% CI rather than SE.*

RESPONSE: Table 2 has been updated to report coefficients and 95% CIs, as have all other tables displaying regression coefficients.