nature portfolio

Peer Review File

Prospective analysis of long-term trajectories of functional scores before and after a diagnosis of stroke

Corresponding Author: Dr Xianwen Shang

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

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
Thank you to the Authors for this work. I enjoyed reading the manuscript, receiving the insights from the data presented, and noting that the work, appropriately, was able to raise further important questions that facilitate further research.

This authors look at four recognised, long established, and extensive public surveys / data sets, to attempt to clarify some of the key bio-psycho-social consequences of stroke, and also using the same data sets to explore PREceding trends and trajectories of the same key elements. This has been done well, and I believe presents new information with regard to a range of bio-psycho-social impacts both before and after a stroke event, and a comparison with an 'equivalent' non-stroke population. I like the choice of Variables selected by the Authors, and I like the choice of Statistical analysis, using Linear Mixed Models, and the process for selecting the relevant number of knots. I also very much appreciated the insights from the Gender based analyses.

1/ Choice of Data-sets is valid, and represents a cross-section of cultures / ethnicities, as the authors have explained. This is very useful in understanding commonalities across these borders, but adds to the complexity of data presentation / explanation. The authors have done well in trying to present the different outcomes from the different data sets (especially with the well-presented Figures), but tying it all together in a common conclusion is not easy. Noting that some of the trajectories appear to significantly differ, offering the Authors opinion and rationale about the 'confidence' of the accuracy and 'meaningfulness' of this 'summation' would be helpful as part of the Discussion.

2/ As noted above, there are a range of 'commonalities' across the trajectories in these data sets, but also some differences. In this paper, the focus is on conclusions from the 'summation' of the data. This, in effect, provides nice information applicable to an 'International' audience, but loses confidence in the Conclusions for some (individual) National audiences. This (former) may be the goal of the authors, but some discussion that contains both comparison as well as contrasts of the trajectories (between nations) would add to the insights for the Discussion, and Conclusions.

3/ Unavoidably, there will be a survivor bias in this type of study. The authors recognize this to some extent in their discussion, but I think that more exploration of the impact of this bias for each of the 'Stroke' and 'non-stroke' trajectories is important, as this would definitely impact on the trajectories of each of the variables measured. Insight into the number of Deaths / dropouts would aid in the interpretation of the trajectories.

4/ Use of a Frailty Index is an important measure, and I am pleased that the Authors have included this. The Frailty Index selected by the authors overlaps significantly with other variables, but also has s stronger bias towards the 'physical' functions (ADLs and Chronic Conditions) (as distinct from Cognition or mood, for example). As a result, it can perhaps be expected that there will be some 'correlation' between (I)ADL outcomes and Frailty outcomes more than the non-physical variables. The authors touch on this, but I think it deserves a small amount of added exploration in Discussion, and elsewhere, especially when considering the associations between this FI and the Trajectories before Stroke, and the Stroke incidence risk.

5/ The authors mention that they created 2 models, unadjusted, and adjusted (for a number of variables). I assume that the analysis is based on the adjusted model, but I am not sure if this was clearly stated?

6/ The median age of the populations studied was a bit lower than I would have expected for Stroke. Do the authors have an explanation for this?, And do they think it impacts on the identified trajectories?

7/ Please note that the Sentence spanning Lines 276-278 is not a proper sentence. Also, spelling error, Line 289, 'whist' = 'whilst'. I also do not understand the phrase '... provides evidence on the critical periods for minimizing memory loss...'(lines 297 - 298). Also, lines 320 - 321, 'the importance of the management of stroke on the first five years after stroke' requires better explanation.

8/ There are parts where the authors appear to attribute possible causality, when the evidence may suggest only an association. One example is in line 311, where there is a comment that 'the steeper increase in CESD' identifies a phase which is important for 'prevention' of stroke, implying Depression causes stroke? Lines 297 - 298 MIGHT have a similar implication? Caution not to create phrases which may imply 'causality' rather than just 'association/predictors'.

Overall, I valued the opportunity to read this manuscript. The authors have used valuable data to explore variables and outcomes to identify trajectories over a longer period of time including pre- and post- stroke. The information presented provides value when considering the epidemiology of stroke as a public health issue, as well as costs to communities as a result of stroke.

Reviewer #2

(Remarks to the Author)
Thank you for the opportunity to review this manuscript. The study addresses an important topic in rehabilitation and epidemiology. I have several methodological concerns and suggestions that would strengthen the manuscript's clarity and rigor.
The manuscript inadequately contextualizes the current study within the existing literature. References 10-12 are cited in the introduction but lack detailed description of their findings. Please provide a comprehensive summary of what each of these previous studies investigated and their key results. This will help readers understand the specific knowledge gap your study addresses. Furthermore, the discussion section notably lacks any comparison between your findings and these previous studies. A thorough comparison with prior literature is essential for interpreting your results and understanding their contribution to the field.
I recommend that the authors provide detailed information on participant attrition, including the number of participants lost to follow-up or deceased at each wave. The follow-up rates for each survey wave should be clearly presented, as this information is crucial for assessing potential attrition bias. Additionally, the definition of baseline requires clarification. Specifically, it is unclear whether baseline in the analysis refers to the time of the first survey following stroke onset, or whether year 0 represents the actual time of stroke onset. This distinction is critical for interpreting the temporal relationships in your findings. The manuscript would also benefit from a more comprehensive explanation of how physical activity was measured and operationalized in the analysis.
The reliance on self-reported physician diagnosis without validation represents a significant methodological limitation. I am wondering if self-reported stroke is validated. This limitation should be thoroughly addressed in the discussion section, including an explicit acknowledgment of this measurement error, discussion of whether this misclassification would produce bias, and if so, the likely direction of such bias (toward or away from the null) as a limitation. Recognition that crucial clinical information—including stroke severity, type (ischemic vs. hemorrhagic), and anatomical location—remains completely unknown in your dataset should also be acknowledged.
The analytical approach requires substantial clarification. The rationale for employing three different modeling approaches (linear mixed model, GEE, and linear regression) is unclear. Furthermore, the figure legend mentions a fourth approach—latent-class mixed models—which is not discussed in the methods. I recommend using a single, appropriate model such as linear mixed models, which can estimate differences between stroke status at each time point and can also estimate annual changes (e.g., from year -4 to year 3) using marginal means. The notation "*refers to significance for linear relationships. † refers to significance for non-linear relationships" requires clarification. Please explain how these different significance tests were conducted and their interpretation. Please specify how random effects were handled—were random slopes, random intercepts, or both included? The mention of "knots" suggests the use of spline curves, but this is not explicitly stated. Please clarify whether splines were used and provide details about their implementation. Model fit statistics (AIC and BIC) should be reported to justify model selection. The distinction between the stroke patient trajectories presented in Figures 1-4 versus those in Figure 5 is unclear. Please provide a detailed explanation in the statistical analysis section regarding what different analytical approaches or data subsets these figures represent.
Please provide detailed information about missing data patterns, including the proportion of missing values for each variable. As multiple imputation requires specific assumptions about the missing data mechanism (MCAR, MAR, or MNAR), these assumptions and their justification should be clearly stated. The approach to testing effect modification by sex appears to involve multiple tests on the same outcome, which raises concerns about multiple testing. Please reconsider this analytical strategy and consider appropriate adjustments for multiple comparisons.
Table 1 should include the sample size for each group to provide readers with a clear understanding of the study population distribution.
These revisions are essential for ensuring the methodological rigor and interpretability of your findings. I look forward to reviewing the revised manuscript with these improvements incorporated.

Reviewer #3

(Remarks to the Author)

I have read carefully the manuscript “Long-term trajectories of memory, depressive symptoms, mobility, and frailty index before and after the diagnosis of stroke: a prospective analysis.”
At first glance the paper appears technically impressive, because it merges data from four of the largest international aging cohorts (CHARLS, ELSA, HRS and SHARE), and it tries to use a kind of harmonized longitudinal approach to study how memory, depression, physical function and frailty evolve before and after a stroke event. This is clearly an important topic: the decline in function before stroke is a critical issue in public health and also for rehabilitation planning, and the idea of aligning trajectories around the index event is interesting. However, after a detailed reading of both the main text and supplementary material, I think the manuscript is not yet mature for publication, and unfortunately I have to recommend rejection at this stage. The paper has several strengths but the weaknesses are numerous and quite structural.

1) Introduction:
-The conceptual rationale is weakly developed (and results substantially confirm rather than updating previous evidence). The introduction lists global stroke statistics, which are fine, but the authors never really explain why we should model trajectories before and after stroke, except for vague phrases like “informing prevention and management.” What kind of prevention? What kind of intervention window? Are they testing a hypothesis that functional decline accelerates due to subclinical vascular burden, or are they just describing patterns? It is unclear. This conceptual vagueness carries through the entire paper: it reads more like a long descriptive report than a mechanistic or hypothesis-driven analysis.
2) Methods:
- Stroke ascertainment is entirely self-reported, which is a very serious limitation not discussed properly. We know from validation studies that self-report in surveys can misclassify up to 30% of stroke events, particularly silent or mild ones. (https://doi.org/10.1371/journal.pone.0137538). Moreover, no distinction is made between first and recurrent strokes, or between ischemic and hemorrhagic types. Stroke severity is unknown, which in reality determines the shape of recovery trajectories. Without those, the findings become blurred averages with high heterogeneity. The authors just treat “stroke yes/no” as a homogeneous exposure, which is unrealistic.
3) Statistical analysis:
-I suggest to briefly discuss why linear mixed models and GEE were both used for the analysis.
-I have some concerns about outcome scaling: according to the methods section, “All cognitive and functional scores were normalized between 0 and 1.” That means that in each cohort, for every outcome (memory, CES-D, ADL, IADL, frailty index), the observed minimum value was mapped to 0 and the maximum to 1; authors then used these normalized scores as dependent variables in mixed models estimating trajectories over time and differences between stroke and non-stroke participants. At first glance, normalization seems convenient as it allows direct comparison of slopes across outcomes that originally have different ranges (e.g., memory 0–20, CES-D 0–30, ADL 0–10) and graphically, all outcomes fit within the same plot range. So it simplifies presentation, but this simplicity comes at a big cost. After rescaling, a slope of –0.01 per year in “normalized memory” has no intuitive meaning. We can’t tell if that corresponds to losing one recalled word, half a word, or one point on a 10-item scale.
Readers and especially clinicians cannot translate the effect into a real-world consequence.
In contrast, saying “memory declined by 0.4 words per year” is concrete and understandable.
So, the standardized βs (often reported with five decimal places) look precise, but they’re not interpretable. They express a fraction of an abstract range, not a tangible change in function.
Furthermore, each cohort has different natural ranges and distributions, so I think that rescaling to 0-1 can lead to loss of information about variance. A “0.1” increase may correspond to very different absolute changes or clinical severity across cohorts.
Finally, min-max scaling depends entirely on the observed minimum and maximum. If one participant scores unusually high or low, it can compress the rest of the scale. In ageing cohorts with skewed distributions (many participants near “no deficit”), this can make moderate changes appear exaggerated.
To adress all these issues, I suggest to consider z-scoring within cohort (subtracting mean and dividing by SD) that preserves comparability in terms of standard deviations rather than arbitrary minima and maxima, and it allows pooled meta-analytic summaries.
-Some overlaps between frailty index and study outcomes (e.g. depression, cognitive items) may inflate associations. Consider sensitivity analyses made by removing overlapping domains.
-Attrition bias is a significant issue in longitudinal studies of aging but it is not appropriately described (apart from study limitations) and addressed: the finding that "memory decline stabilizes after several years post-stroke” and “ADL differences remain stable or even decrease after 5 years.” may be more a result of attrition than of recovery or stabilization. I suggest to include at least weighting by inverse probability of follow-up (considering lost to FU as missing not at random).
This is a relevant point as one of the novel results of the paper is related.
-Imputation model: The use of “fully conditional specification” with 10 imputations is reasonable but too briefly justified; specify which variables were imputed, proportion of missingness, and diagnostics performed.
-Definition and placement of knots, AIC adn BIC should be improved. Authors mention linear mixed models with knots selected by AIC and BIC but never show the exact model form, random effects, or knot placement. It’s unclear whether they allowed random slopes, or how the number of observations per participant was handled.
-From the methods section, the manuscript never mentions the use of survey or longitudinal weights included in the framework of longitudinal studies of aging. Weights cannot be omitted as the study’s objective is to describe population-level trajectories before and after stroke, so they become essential. Otherwise, the “average trajectory” reported corresponds only to the analytic subset, not to the true national aging population.
4) Results and discussion:
The results themselves are consistent with prior knowledge: cognitive decline and frailty increase before stroke, and ADL limitations worsen after. So it seems that this study is more confirmatory than novel. However, after carefully reading the discussion, there are some novel aspects that deserve to be better addressed to verify their consistency:
-"We also found the difference in ADL and IADL between stroke and non-stroke participants remained stable or even

decreased 5 years after stroke": this deserve to be verified after appropriately addressing attrition bias.
-Gender-related differences are simply reported and not tried to be justified (by a biological viewpoint); my question is why do memory scores decline faster in men but frailty increases more in women? Is it biological or social?
-Consider synthesis of most relevant results by using pooled estimates or meta-analysis to allow a better visualization and comparison of study results across the cohorts.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
I have read the revised paper, the comments of the other reviewers, and the rebuttal responses from the Authors. I particularly appreciated the comments from the other reviewers.

I believe that the Authors have addressed my points reasonably. I believe that this paper highlights potentially important conclusions when looking at trajectories (both pre-stroke as well as post-stroke) of various health related functions in the cohorts examined.

Are the Conclusions valid and credible, based on the information provided? I think that the credibility of the outcomes has been improved significantly following the revisions based on the comments/recommendations of the other reviewers. Their insights may result in further revisions of the paper, but for me, the paper, as it now stands, has enough validity to stimulate further valuable research based on the Conclusions outlined in the paper by these authors, if the paper is published. Some questions about the accuracy of the data, and its interpretation will remain, and I also continue to have some concerns about its generalisability, for example, but I think that it is up to others to refine and improve the research in this arena. This paper, if published, I believe will allow others to 'build a better study'; in the meantime, the conclusions reached by these authors, in this paper, based upon the data they have presented, and their analytical methods, are reasonable to put forward to 'open discussion'(or build on current evidence and interpretation).

I have no further recommendations, but will be interested to see what the other reviewers suggest.

Reviewer #2

(Remarks to the Author)
Thank you for the revision of the manuscript. While some points have been substantially improved, several critical issues remain insufficiently addressed, and the authors have not provided adequate justification for why certain suggested revisions were not incorporated. Therefore, my recommendation is unfortunately to reject the manuscript. I have provided my comments below, which largely correspond to those raised in my previous review.

1. In my prior review, I recommended substantial revisions to the Introduction, as the conceptual framework lacked clarity. The authors have added a summary of previous research. However, my original suggestion was to articulate the evidence gap more explicitly. Despite the added summary, the evidence gap remains unclear, and the motivation for conducting this study is not sufficiently conveyed. I would encourage the authors to further revise the Introduction to clearly delineate the specific gap in the existing literature that this study aims to address.

2. The baseline definition has been clarified, which is appreciated. However, I have concerns regarding the matching process. The control group was matched on baseline age and sex, which raises the possibility that some matched individuals may have died before the allocated stroke incident date. Furthermore, Table S18 indicates that all participants in ELSA, HRS, and SHARE completed at least three assessments, which seems strange. In relation to the matching process, the population inclusion criteria and how they were followed remain unclear. As currently described, the matching process and population selection are either unreliable or insufficiently explained.

3. The authors have now presented mortality and the number of assessments across groups, which reveals considerable difference. I originally suggested reporting this information because such imbalances could introduce bias or affect estimation. The authors discussed that attrition bias was unlikely, citing consistent results from the imputation analysis. However, as I elaborate in the following comment, it is unclear whether the imputation itself was appropriately conducted. This point therefore warrants further discussion.

4. The assumed missing data mechanism has not been described. This is an essential consideration for the validity of multiple imputation of the outcome, which relies on specific assumptions to produce unbiased estimates. Without this information, it is not possible to evaluate whether the imputation approach was appropriate. (10.1093/ije/dyz032)

5. The modelling strategy remains unclear. The authors appear to have both linear versus non-linear relationships and subsequently employed restricted cubic splines. I was not able to understand why the authors used multiple trajectory description. The rationale for combining both methods is not explained, and readers cannot readily follow the analytical procedures that were conducted.

Reviewer #3

(Remarks to the Author)
The authors significantly improve the quality of results and the way thwy are presented. I have only some minor comments to report:

-Revise some expressions and soften the tone of some sentences: e.g. "The observation that ADL and IADL trajectories begin to diverge several years before a stroke... suggests two intervention windows." Evaluate to replace with "two potential intervention windows". Change "there may be a bidirectional relationships" with "there may be a bidirectional relationship".
-The phrase "Several years before and after" appears frequently. Considering replacing some with "“in the peri-stroke period”, “around the time of stroke”, “in the years surrounding stroke”.

Version 2:

Reviewer comments:

Reviewer #2

(Remarks to the Author)
Thank you for the revision of the manuscript. I think the manuscript has been much improved. Therefore, I recommend an accept decision.

Reviewers' comments:
Reviewer #1 (Remarks to the Author):

Thank you to the Authors for this work. I enjoyed reading the manuscript, receiving the insights from the data presented, and noting that the work, appropriately, was able to raise further important questions that facilitate further research.

This authors look at four recognised, long established, and extensive public surveys / data sets, to attempt to clarify some of the key bio-psycho-social consequences of stroke, and also using the same data sets to explore PREceding trends and trajectories of the same key elements. This has been done well, and I believe presents new information with regard to a range of bio-psycho-social impacts both before and after a stroke event, and a comparison with an 'equivalent' non-stroke population. I like the choice of Variables selected by the Authors, and I like the choice of Statistical analysis, using Linear Mixed Models, and the process for selecting the relevant number of knots. I also very much appreciated the insights from the Gender based analyses.

Response:

We thank the Reviewer for the positive comments.

1/ Choice of Data-sets is valid, and represents a cross-section of cultures / ethnicities, as the authors have explained. This is very useful in understanding commonalities across these borders, but adds to the complexity of data presentation / explanation. The authors have done well in trying to present the different outcomes from the different data sets (especially with the well-presented Figures), but tying it all together in a common conclusion is not easy. Noting that some of the trajectories appear to significantly differ, offering the Authors opinion and rationale about the 'confidence' of the accuracy and 'meaningfulness' of this 'summation' would be helpful as part of the Discussion.

Response:

We agree with the Reviewer that some of the trajectories (memory and frailty index) appear to significantly differ between cohorts, although trajectories of depressive symptoms and ADL were similar across cohorts. We have mentioned this as a limitation in the Discussion section:

Fourth, the trajectories of certain scores (memory and frailty) in CHARLS and SHARE differed from those in ELSA and HRS, while ELSA and HRS exhibited similar trajectories across all scores. These discrepancies may reflect differences in age, ethnicity, and follow-up duration across cohorts.

2/ As noted above, there are a range of 'commonalities' across the trajectories in these data sets, but also some differences. In this paper, the focus is on conclusions from the 'summation' of the data. This, in effect, provides nice information applicable to an 'International' audience, but loses confidence in the Conclusions for some (individual) National audiences. This (former) may be the goal of the authors, but some discussion that contains both comparison as well as contrasts of the trajectories (between nations) would add to the insights for the Discussion, and Conclusions.

Response:

We appreciate the suggestion to balance international summation with cohort-specific contrasts.

We have expanded the Discussion and Conclusions to interpret similarities and differences across national contexts*:*

Fourth, the trajectories of certain scores (memory and frailty) in CHARLS and SHARE differed from those in ELSA and HRS, while ELSA and HRS exhibited similar trajectories across all scores. These discrepancies may reflect differences in age, ethnicity, and follow-up duration across cohorts.

While the overall international patterns were consistent, cohort-specific contrasts in the magnitude and shape of trajectories suggest that national context—measurement, care pathways, and social supports—may modulate functional changes around stroke.

3/ Unavoidably, there will be a survivor bias in this type of study. The authors recognize this to some extent in their discussion, but I think that more exploration of the impact of this bias for each of the 'Stroke' and 'non-stroke' trajectories is important, as this would definitely impact on the trajectories of each of the variables measured. Insight into the number of Deaths / dropouts would aid in the interpretation of the trajectories.

Response:

We agree with the Reviewer that survivor bias could influence the observed functional trajectories. To address this, we have now provided the incidence of mortality for individuals with and without incident stroke in Table 1. As our analysis included only participants with at least two waves of data, there is no attrition (i.e., loss to follow-up while alive) within the analytical sample. However, to provide a complete picture of data availability, we have added the frequency distribution for the number of surveys completed per participant.

4/ Use of a Frailty Index is an important measure, and I am pleased that the Authors have included this. The Frailty Index selected by the authors overlaps significantly with other variables, but also has s stronger bias towards the 'physical' functions (ADLs and Chronic Conditions) (as distinct from Cognition or mood, for example). As a result, it can perhaps be expected that there will be some 'correlation' between (I)ADL outcomes and Frailty outcomes more than the non-physical variables. The authors touch on this, but I think it deserves a small amount of added exploration in Discussion, and elsewhere, especially when considering the associations between this FI and the Trajectories before Stroke, and the Stroke incidence risk.

Response:

We have some discussion about this:

The trajectories of the frailty index over the years following a stroke are similar to those of ADL/IADL in our study. This may be attributed to the fact that the calculation of the frailty index involves components such as memory, depressive symptoms, and ADL/IADL, with ADL/IADL accounting for a significant weight.

5/ The authors mention that they created 2 models, unadjusted, and adjusted (for a number of variables). I assume that the analysis is based on the adjusted model, but I am not sure if this was clearly stated?

Response:

Thank the Reviewer for raising this concern. The results are consistent between unadjusted and

adjusted models, so we just reported the adjusted results. We have noted whether the reported results are unadjusted or adjusted.

6/ The median age of the populations studied was a bit lower than I would have expected for Stroke. Do the authors have an explanation for this?, And do they think it impacts on the identified trajectories?

Response:

We agree with the Reviewer that the median age of the populations studied was a bit lower than expected, but this may not highly impact the identified trajectories. We have added discussion about this:

Fifth, the median age of included stroke participants was somewhat lower than that of typical stroke cohorts, which may be due to stroke patients being more likely to be lost to follow-up because of higher mortality risk (only those with two or more surveys were included in the analysis). It may also be attributable to the increasing incidence of stroke among young adults in recent decades globally.[31] However, the lower median age is unlikely to substantially influence the identified trajectory patterns, especially given age matching and age adjustment.

7/ Please note that the Sentence spanning Lines 276-278 is not a proper sentence. Also, spelling error, Line 289, 'whist' = 'whilst'. I also do not understand the phrase '... provides evidence on the critical periods for minimizing memory loss...'(lines 297 - 298). Also, lines 320 - 321, 'the importance of the management of stroke on the first five years after stroke' requires better explanation.

Response:

We thank the Reviewer for raising the grammar errors in the manuscript. We have revised the sentences accordingly.

8/ There are parts where the authors appear to attribute possible causality, when the evidence may suggest only an association. One example is in line 311, where there is a comment that 'the steeper increase in CESD' identifies a phase which is important for 'prevention' of stroke, implying Depression causes stroke? Lines 297 - 298 MIGHT have a similar implication? Caution not to create phrases which may imply 'causality' rather than just 'association/predictors'.

Response:

We have revised the sentences to not imply causality for our findings.

Overall, I valued the opportunity to read this manuscript. The authors have used valuable data to explore variables and outcomes to identify trajectories over a longer period of time including pre- and post-stroke. The information presented provides value when considering the epidemiology of stroke as a public health issue, as well as costs to communities as a result of stroke.

Response:

We thank the Reviewer for the positive comments and we have revised the manuscript accordingly.

Reviewer #2 (Remarks to the Author):

Thank you for the opportunity to review this manuscript. The study addresses an important topic in rehabilitation and epidemiology. I have several methodological concerns and suggestions that would strengthen the manuscript's clarity and rigor.

Response:

We thank the Reviewer for the insight comments and we have revised the manuscript accordingly.

The manuscript inadequately contextualizes the current study within the existing literature. References 10-12 are cited in the introduction but lack detailed description of their findings. Please provide a comprehensive summary of what each of these previous studies investigated and their key results. This will help readers understand the specific knowledge gap your study addresses. Furthermore, the discussion section notably lacks any comparison between your findings and these previous studies. A thorough comparison with prior literature is essential for interpreting your results and understanding their contribution to the field.

Response:

We have added the main findings of these individual studies.

The references raising by the Reviewer are related to changes in functional scores poststroke. Our study is more relevant to changes in functional scores before and after stroke. Such that we included a lot of comparisons between my findings and previous studies regarding changes in functional scores before and after stroke. For example, we have discussion section regarding the comparison between my findings and these previous studies:

Our study is consistent with previous research showing a significant decline in memory among both stroke and non-stroke participants,[12,18,27] with a steeper decline observed in stroke patients. A meta-analysis also demonstrated that stroke patients experienced higher cognitive decline than that of non-stroke controls from 1 to 3 years after onset.[12] In contrast, data from the SHARE with 2,362 incident strokes between 2004 and 2019 revealed an acute decrease in memory after stroke but similar decline between stroke and non-stroke participants.[14]

I recommend that the authors provide detailed information on participant attrition, including the number of participants lost to follow-up or deceased at each wave. The follow-up rates for each survey wave should be clearly presented, as this information is crucial for assessing potential attrition bias. Additionally, the definition of baseline requires clarification. Specifically, it is unclear whether baseline in the analysis refers to the time of the first survey following stroke onset, or whether year 0 represents the actual time of stroke onset. This distinction is critical for interpreting the temporal relationships in your findings. The manuscript would also benefit from a more comprehensive explanation of how physical activity was measured and operationalized in the analysis.

Response:

We thank the Reviewer for the valuable comments. We have added the incidence of mortality and the number of participants with different number of surveys in Table 1.

'Baseline' in the analysis refers to the time of the first survey of included participants. We have added this explanation to the notes of Table 1:
Baseline in the analysis refers to the time of the first survey of included participants.

We have also added information regarding how physical activity was assessed. Physical activity was treated as a categorical variable.

The reliance on self-reported physician diagnosis without validation represents a significant methodological limitation. I am wondering if self-reported stroke is validated. This limitation should be thoroughly addressed in the discussion section, including an explicit acknowledgment of this measurement error, discussion of whether this misclassification would produce bias, and if so, the likely direction of such bias (toward or away from the null) as a limitation. Recognition that crucial clinical information—including stroke severity, type (ischemic vs. hemorrhagic), and anatomical location—remains completely unknown in your dataset should also be acknowledged.

Response:

We agree with the Reviewer that the definition of stroke is limited by self-reported physician diagnosis only. Stroke was ascertained in two steps. First, stroke was confirmed if the year of diagnosis or age at diagnosis was reported. Second, stroke was confirmed if it was reported again in a subsequent survey wave.

The misclassifications were more likely to be false positives, so that the associations between stroke and trajectories of functional scores were more likely to be toward to the null. The stroke severity, type (ischemic vs. hemorrhagic), and anatomical location are unknown in the current analysis, which might have influence the associations. We have added discussion about this:
Although stroke was confirmed if the year or age at diagnosis was reported, or if it was reported again in a subsequent survey wave, reliance on self-reported physician diagnosis may have resulted in measurement error. Misclassification was more likely to involve false positives, which would tend to bias the associations between stroke and trajectories of functional scores toward the null. In addition, stroke severity, type, and anatomical location were unknown in the current analysis, which may have influenced the observed associations.

The analytical approach requires substantial clarification. The rationale for employing three different modeling approaches (linear mixed model, GEE, and linear regression) is unclear. Furthermore, the figure legend mentions a fourth approach—latent-class mixed models—which is not discussed in the methods. I recommend using a single, appropriate model such as linear mixed models, which can estimate differences between stroke status at each time point and can also estimate annual changes (e.g., from year -4 to year 3) using marginal means. The notation "*refers to significance for linear relationships. † refers to significance for non-linear relationships" requires clarification. Please explain how these different significance tests were conducted and their interpretation. Please specify how random effects were handled—were random slopes, random intercepts, or both included? The mention of "knots" suggests the use of spline curves, but this is not explicitly stated. Please clarify whether splines were used and provide details about their implementation. Model fit statistics (AIC and BIC) should be reported to justify model selection. The distinction between the stroke patient trajectories

presented in Figures 1-4 versus those in Figure 5 is unclear. Please provide a detailed explanation in the statistical analysis section regarding what different analytical approaches or data subsets these figures represent.

Response:

We thank the Reviewer for raising this important concern. As the Reviewer suggests, we have used linear mixed models as the single method to estimate differences between stroke status at each time point and the annual changes over time. "latent-class mixed models" is a typo error and we have removed the sentence.

We have added the explanation for linear and non-linear relationships in the methods section and the notes of the figures:

Both linear and non-linear relationships were tested. Linear relationships were tested using the fixed effect coefficient in a linear mixed model. Non-linear relationships were tested by comparing models with and without quadratic/spline terms using likelihood ratio tests.

For random effects, we have included both intercepts and slopes as random effects. We have added more information about this:

Within-participant correlations were accounted for by correlated random intercepts and slopes of time and time squares. Splines were integrated into the models to capture potential variations in trajectories of individual scores over time.

For model selection, we have added Table S14-S17 for AIC and BIC statistics. Figures 1-4 provide the trajectories of individual scores for those with and without stroke, which aim to compare the difference in the changes in scores between stroke and non-stroke participants. While Figure 5 aims to compare the changes between scores among stroke patients.

Please provide detailed information about missing data patterns, including the proportion of missing values for each variable. As multiple imputation requires specific assumptions about the missing data mechanism (MCAR, MAR, or MNAR), these assumptions and their justification should be clearly stated. The approach to testing effect modification by sex appears to involve multiple tests on the same outcome, which raises concerns about multiple testing. Please reconsider this analytical strategy and consider appropriate adjustments for multiple comparisons.

Response:

We have added more information on the imputation model:

Multiple imputations for missing data in continuous variables were conducted for outcomes and covariates using the fully conditional method. This method imputes each variable with missing values using a regression model conditional on the other variables, iterating through all variables until convergence. It is flexible and easy to implement, allowing variable-wise models and accommodating different types of variables. All covariates were included in the imputation models to create 10 imputed datasets.

We also agree with the Reviewer that appropriate methods should be used to deal with multiple

comparisons. We have added the information in the methods section and updated the results accordingly (Figures S33-S36, S38-S41, S43-S46, S48-S51, and S53-S56).
Given the multiple comparisons, false discovery rate–adjusted P-values for the interaction by gender were reported.

Table 1 should include the sample size for each group to provide readers with a clear understanding of the study population distribution.

Response:

We have included the sample size for each group in Table 1.

These revisions are essential for ensuring the methodological rigor and interpretability of your findings. I look forward to reviewing the revised manuscript with these improvements incorporated.

Response:

We thank the Reviewer for the valuable comments. We believe our manuscript has been improved much after the revisions.

Reviewer #3 (Remarks to the Author):

I have read carefully the manuscript "Long-term trajectories of memory, depressive symptoms, mobility, and frailty index before and after the diagnosis of stroke: a prospective analysis."

At first glance the paper appears technically impressive, because it merges data from four of the largest international aging cohorts (CHARLS, ELSA, HRS and SHARE), and it tries to use a kind of harmonized longitudinal approach to study how memory, depression, physical function and frailty evolve before and after a stroke event. This is clearly an important topic: the decline in function before stroke is a critical issue in public health and also for rehabilitation planning, and the idea of aligning trajectories around the index event is interesting. However, after a detailed reading of both the main text and supplementary material, I think the manuscript is not yet mature for publication, and unfortunately I have to recommend rejection at this stage. The paper has several strengths but the weaknesses are numerous and quite structural.

Response:

We thank the Reviewer for their positive and constructive comments. We agree with the points raised regarding the limitations of our study and have revised the manuscript accordingly.

1) Introduction:

-The conceptual rationale is weakly developed (and results substantially confirm rather than updating previous evidence). The introduction lists global stroke statistics, which are fine, but the authors never really explain why we should model trajectories before and after stroke, except for vague phrases like "informing prevention and management." What kind of prevention? What kind of intervention window? Are they testing a hypothesis that functional decline accelerates due to subclinical vascular burden, or are they just describing patterns? It is unclear. This conceptual vagueness carries through the entire paper: it reads more like a long descriptive report than a mechanistic or hypothesis-driven analysis.

Response:

We thank the Reviewer for this insightful comment. We agree that our original Introduction did not sufficiently articulate the conceptual framework and hypotheses guiding the analysis. Accordingly, we have substantially revised the Introduction to address this point.

We have also revised the Discussion to more explicitly connect our empirical findings to the stated hypotheses and to clarify how the identified trajectories inform specific strategies for prevention and post-stroke management.

2) Methods:

- Stroke ascertainment is entirely self-reported, which is a very serious limitation not discussed properly. We know from validation studies that self-report in surveys can misclassify up to 30% of stroke events, particularly silent or mild ones. (https://doi.org/10.1371/journal.pone.0137538). Moreover, no distinction is made between first and recurrent strokes, or between ischemic and hemorrhagic types. Stroke severity is unknown, which in reality determines the shape of recovery trajectories. Without those, the findings become blurred averages with high heterogeneity. The authors just treat "stroke yes/no" as a homogeneous exposure, which is unrealistic.

Response:

Response:

We agree with the Reviewer that the definition of stroke is limited by self-reported physician diagnosis only. Stroke was ascertained in two steps. First, stroke was confirmed if the year of diagnosis or age at diagnosis was reported. Second, stroke was confirmed if it was reported again in a subsequent survey wave.

The stroke severity, recurrence, type (ischemic vs. hemorrhagic), and anatomical location are not tested in the current analysis, which might have influence the associations. We have added discussion about this:

Although stroke was confirmed if the year or age at diagnosis was reported, or if it was reported again in a subsequent survey wave, reliance on self-reported physician diagnosis may have resulted in measurement error. Misclassification was more likely to involve false positives, which would tend to bias the associations between stroke and trajectories of functional scores toward the null. In addition, stroke severity, type, and anatomical location were unknown in the current analysis, which may have influenced the observed associations.

3) Statistical analysis:

-I suggest to briefly discuss why linear mixed models and GEE were both used for the analysis.

Response:

Both linear mixed models and GEE are appropriate for testing the longitudinal changes over time. To make it consistency, we have replaced GEE with linear mixed methods in the analysis.

All cognitive and functional scores were normalized between 0 and 1. Linear mixed models were used to estimate the trajectories of each score over time in individuals with and without incident stroke. The optimal number of knots was determined using the Akaike and Bayesian information criteria. The annual changes in each score for each time period between knots were analyzed using linear mixed models. Two models were tested: 1) unadjusted (model 1); 2) adjusted for age, gender, education, smoking, physical activity, alcohol consumption, BMI, hypertension, and diabetes (model 2). The mean difference in individual scores between those with and without incident stroke for each two-year period from stroke diagnosis was examined using linear mixed models.

-I have some concerns about outcome scaling: according to the methods section, “All cognitive and functional scores were normalized between 0 and 1.” That means that in each cohort, for every outcome (memory, CES-D, ADL, IADL, frailty index), the observed minimum value was mapped to 0 and the maximum to 1; authors then used these normalized scores as dependent variables in mixed models estimating trajectories over time and differences between stroke and non-stroke participants. At first glance, normalization seems convenient as it allows direct comparison of slopes across outcomes that originally have different ranges (e.g., memory 0–20, CES-D 0–30, ADL 0–10) and graphically, all outcomes fit within the same plot range. So it simplifies presentation, but this simplicity comes at a big cost. After rescaling, a slope of −0.01 per year in “normalized memory” has no intuitive meaning. We can’t tell if that corresponds to losing one recalled word, half a word, or one point on a 10-item scale.

Response:

We agree with the Reviewer that normalizing scores facilitates the direct comparison of

trajectories across different outcomes. To implement this while preserving interpretability, we have revised the analysis to use z-scores (calculated as [value - mean] / SD) as the unit of measurement. This approach standardizes the scales, enabling slope comparisons, while the z-score unit itself remains statistically interpretable. We have updated the results and figures accordingly.

Readers and especially clinicians cannot translate the effect into a real-world consequence.
In contrast, saying "memory declined by 0.4 words per year" is concrete and understandable.
So, the standardized βs (often reported with five decimal places) look precise, but they're not interpretable. They express a fraction of an abstract range, not a tangible change in function.

Response:

Because the original individual scores range from 0 to 1, the estimated coefficients (βs) were very small, requiring five decimal places to discern differences. To improve interpretability, we repeated the analyses using z-scores. This standardization allows results to be meaningfully reported with two or three decimal places. All results have been updated accordingly.

Furthermore, each cohort has different natural ranges and distributions, so I think that rescaling to 0-1 can lead to loss of information about variance. A "0.1" increase may correspond to very different absolute changes or clinical severity across cohorts.

Response:

We totally agree with the Reviewer that rescaling to 0-1 may result in loss of information on variance. Z-scores have been used instead in the analysis.

Finally, min-max scaling depends entirely on the observed minimum and maximum. If one participant scores unusually high or low, it can compress the rest of the scale. In ageing cohorts with skewed distributions (many participants near "no deficit"), this can make moderate changes appear exaggerated.

Response:

Z-scores have been used instead in the analysis.

To adress all these issues, I suggest to consider z-scoring within cohort (subtracting mean and dividing by SD) that preserves comparability in terms of standard deviations rather than arbitrary minima and maxima, and it allows pooled meta-analytic summaries.

Response:

As the Reviewer suggests, we have used Z-scores instead in the analysis. All the results have been updated in the manuscript.

-Some overlaps between frailty index and study outcomes (e.g. depression, cognitive items) may inflate associations. Consider sensitivity analyses made by removing overlapping domains.

Response:

The frailty index was calculated as a comprehensive measure encompassing comorbidities, functional status, cognition, depression, health attitudes/practices, and physical performance. Removing the domains of cognition, depression, and ADL/IADL would leave a score based primarily on 8 chronic diseases and 3 self-reported health

conditions. This modified score would be a less appropriate measure for analyzing longitudinal changes in overall function, as it would no longer represent the intended multidimensional construct of frailty.

-Attrition bias is a significant issue in longitudinal studies of aging but it is not appropriately described (apart from study limitations) and addressed: the finding that "memory decline stabilizes after several years post-stroke" and "ADL differences remain stable or even decrease after 5 years." may be more a result of attrition than of recovery or stabilization. I suggest to include at least weighting by inverse probability of follow-up (considering lost to FU as missing not at random).

Response:

We agree with the Reviewer that attrition bias is an important consideration. We have now provided the cohort-specific attrition rates. To assess its potential impact, we performed sensitivity analyses using imputed data for the key outcomes (cognition, depression, ADL, IADL, and frailty index). The results, shown in Figures 1-5 (and Figure S20) versus Figures S57-S62, are consistent between the complete-case and imputed analyses, suggesting that attrition bias did not substantially alter our findings.

For example, all individual scores show similar trajectories before (Figure A1) and after imputation (Figure A2).

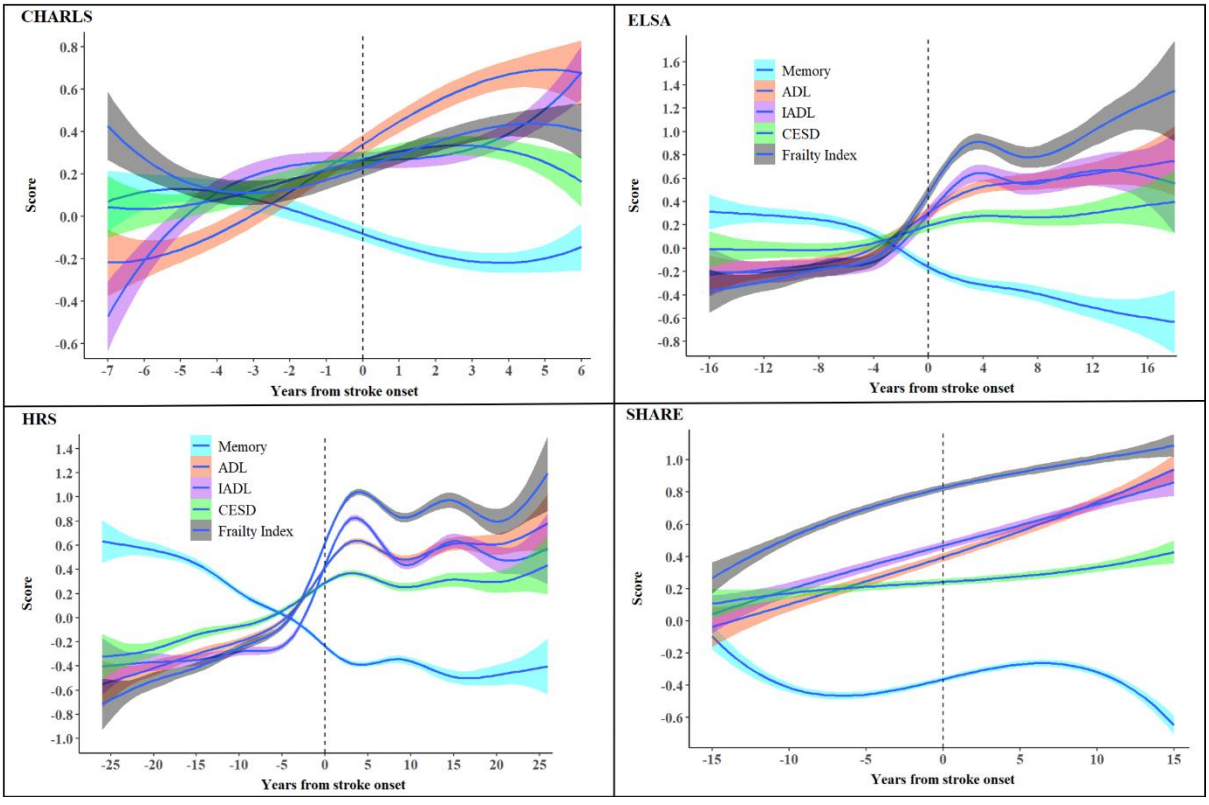

Figure A1. Trajectories of individual scores before and after the diagnosis of stroke among individuals with stroke

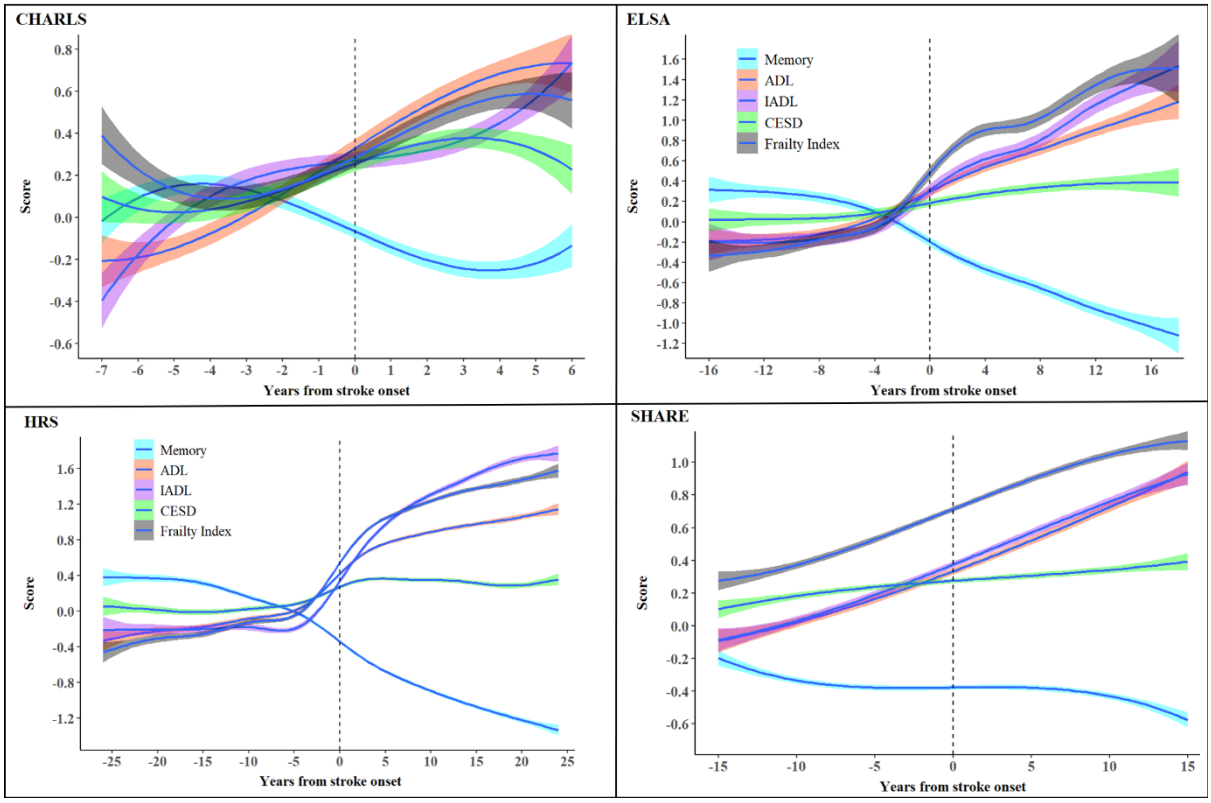

Figure A2. Trajectories of individual scores before and after the diagnosis of stroke among individuals with stroke, using imputed data

This is a relevant point as one of the novel results of the paper is related.
-Imputation model: The use of "fully conditional specification" with 10 imputations is reasonable but too briefly justified; specify which variables were imputed, proportion of missingness, and diagnostics performed.

Response:

We have added more information on the imputation model:

Multiple imputations for missing data in continuous variables were conducted for outcomes and covariates using the fully conditional method. This method imputes each variable with missing values using a regression model conditional on the other variables, iterating through all variables until convergence. It is flexible and easy to implement, allowing variable-wise models and accommodating different types of variables. All covariates were included in the imputation models to create 10 imputed datasets.

The number of participants with missing data is presented in Table 1. The trajectories of individual scores were consistent between the complete-case analysis (Figures 1-5, S20) and the analysis using imputed data (Figures S57-S62).

-Definition and placement of knots, AIC adn BIC should be improved. Authors mention linear mixed models with knots selected by AIC and BIC but never show the exact model form, random effects, or knot placement. It's unclear whether they allowed random slopes, or how the number of observations

per participant was handled.

Response:

The models included both random intercepts and slopes. Specifically, within-participant correlations were accounted for by including correlated random intercepts and slopes for the linear and quadratic terms of time. Flexible trajectories of individual scores over time were modeled using regression splines.

Model fit statistics (AIC and BIC) for all tested models are provided in Tables S14–S17. Sensitivity analyses were performed to ensure robustness: models were run using both complete-case data and imputed data (which included all participants with follow-up records in the same cohort). The results were consistent between these two analytical samples, indicating that the findings are not substantially influenced by the handling of missing data.

-From the methods section, the manuscript never mentions the use of survey or longitudinal weights included in the framework of longitudinal studies of aging. Weights cannot be omitted as the study's objective is to describe population-level trajectories before and after stroke, so they become essential. Otherwise, the "average trajectory" reported corresponds only to the analytic subset, not to the true national aging population.

Response:

We agree with the Reviewer regarding the importance of sampling weights. Accordingly, we have re-run all analyses incorporating the appropriate weights. The Methods section has been updated to state:

Sampling weights were applied to obtain estimates of the differences in individual scores between individuals with and without stroke, as well as the changes in individual scores over time.

We have also updated all results in the manuscript (Figures S5-S19, S21-S31, S33-S36, S38-S41, S43-S46, S48-S51, and S53-S56).

4) Results and discussion:

The results themselves are consistent with prior knowledge: cognitive decline and frailty increase before stroke, and ADL limitations worsen after. So it seems that this study is more confirmatory than novel. However, after carefully reading the discussion, there are some novel aspects that deserve to be better addressed to verify their consistency:

-"We also found the difference in ADL and IADL between stroke and non-stroke participants remained stable or even decreased 5 years after stroke": this deserve to be verified after appropriately addressing attrition bias.

Response:

We conducted a sensitivity analysis using imputed data, which showed trajectories consistent with the complete-case analysis. This indicates that our findings are robust to potential attrition bias. We have added the following to the discussion:

This pattern could potentially be attributed to attrition bias if individuals with more severe

impairment were disproportionately lost to follow-up. However, sensitivity analyses using imputed data produced results consistent with those from the complete-case analysis, suggesting that attrition bias is unlikely to be the primary driver of the observed trajectory.

-Gender-related differences are simply reported and not tried to be justified (by a biological viewpoint); my question is why do memory scores decline faster in men but frailty increases more in women? Is it biological or social?

Response:

The new analysis shows that the memory score decline did not consistently differ between men and women. While CESD, ADL, and IADL appeared to increased faster in women than in men. We have added explanation for this:

Women, on average, experience their first stroke at an older age than men (not shown in our study) and tend to have more severe strokes.[30] These factors may partly explain their worse outcomes following stroke.

-Consider synthesis of most relevant results by using pooled estimates or meta-analysis to allow a better visualization and comparison of study results across the cohorts.

Response:

We thank the Reviewer for raising this point. However, due to differences in follow-up durations and knot placements across studies, pooling estimates for the longitudinal changes in individual scores is methodologically problematic.

Reviewers' comments:
Reviewer #1 (Remarks to the Author):

I have read the revised paper, the comments of the other reviewers, and the rebuttal responses from the Authors. I particularly appreciated the comments from the other reviewers.

I believe that the Authors have addressed my points reasonably. I believe that this paper highlights potentially important conclusions when looking at trajectories (both pre-stroke as well as post-stroke) of various health related functions in the cohorts examined.

Are the Conclusions valid and credible, based on the information provided? I think that the credibility of the outcomes has been improved significantly following the revisions based on the comments/recommendations of the other reviewers. Their insights may result in further revisions of the paper, but for me, the paper, as it now stands, has enough validity to stimulate further valuable research based on the Conclusions outlined in the paper by these authors, if the paper is published. Some questions about the accuracy of the data, and its interpretation will remain, and I also continue to have some concerns about its generalisability, for example, but I think that it is up to others to refine and improve the research in this arena. This paper, if published, I believe will allow others to 'build a better study'; in the meantime, the conclusions reached by these authors, in this paper, based upon the data they have presented, and their analytical methods, are reasonable to put forward to 'open discussion'(or build on current evidence and interpretation).

I have no further recommendations, but will be interested to see what the other reviewers suggest.

Response:

We thank the reviewer for their thoughtful final evaluation and for recognizing the importance of our findings regarding pre- and post-stroke trajectories across these diverse cohorts. We are pleased that the revisions have improved the credibility and validity of the outcomes in the reviewer's estimation.

Reviewer #2 (Remarks to the Author):

Thank you for the revision of the manuscript. While some points have been substantially improved, several critical issues remain insufficiently addressed, and the authors have not provided adequate justification for why certain suggested revisions were not incorporated. Therefore, my recommendation is unfortunately to reject the manuscript. I have provided my comments below, which largely correspond to those raised in my previous review.

1. In my prior review, I recommended substantial revisions to the Introduction, as the conceptual framework lacked clarity. The authors have added a summary of previous research. However, my original suggestion was to articulate the evidence gap more explicitly. Despite the added summary, the evidence gap remains unclear, and the motivation for conducting this study is not sufficiently conveyed. I would encourage the authors to further revise the Introduction to clearly delineate the specific gap in the existing literature that this study aims to address.

Response:

We thank the Reviewer for raising this important concern. We have revised the paragraph of the Introduction section to elucidate the evidence gap of the study.

2. The baseline definition has been clarified, which is appreciated. However, I have concerns regarding the matching process. The control group was matched on baseline age and sex, which raises the possibility that some matched individuals may have died before the allocated stroke incident date. Furthermore, Table S18 indicates that all participants in ELSA, HRS, and SHARE completed at least three assessments, which seems strange. In relation to the matching process, the population inclusion criteria and how they were followed remain unclear. As currently described, the matching process and population selection are either unreliable or insufficiently explained.

Response:

We agree with the Reviewer that some matched individuals may have died before the allocated stroke incident date. Given that controls must be alive and stroke-free at the time their matched case has a stroke. To ensure sufficient data for robust trajectory estimation using linear mixed models, we included only those participants with at least two assessments in CHARLS (out of 5 possible) or at least three assessments in the other cohorts (up to 10 in ELSA, 14 in HRS, and 7 in SHARE). The exclusion and inclusion criteria ensure participants had at least one survey assessment before the stroke onset. As most stroke cases occurred before the last survey assessment, we also have sufficient data for the test of the trajectories after stroke onset. We have added explanation for this in the Methods section:

To ensure valid comparisons, controls were required to be alive and stroke-free at the index date (the time of the matched case's diagnosis). To provide sufficient data for robust trajectory estimation via linear mixed models, we included only those participants with at least two assessments in CHARLS (out of 5 possible) or at least three assessments in the other cohorts (up to 10 in ELSA, 14 in HRS, and 7 in SHARE). This criterion ensures that estimated slopes reflect long-term functional changes rather than random fluctuations between only two time points.

3. The authors have now presented mortality and the number of assessments across groups, which reveals considerable difference. I originally suggested reporting this information because such imbalances could introduce bias or affect estimation. The authors discussed that attrition bias was unlikely, citing consistent results from the imputation analysis. However, as I elaborate in the following comment, it is unclear whether the imputation itself was appropriately conducted. This point therefore warrants further discussion.

Response:

We thank the Reviewer for raising this concern and we have added the explanation as a limitation:

We also observed expected differences in mortality and assessment frequency between cases and controls. Such imbalances can introduce attrition bias, where 'healthier' survivors are over-represented in later years. We addressed this by utilizing linear mixed models, which are robust to data missing at random, and by performing multiple imputation sensitivity analyses that included mortality as a predictor. The consistency between these results and our primary analysis suggests that the fundamental shape of functional decline is not solely driven by the loss of the most severely affected participants.

4. The assumed missing data mechanism has not been described. This is an essential consideration for the validity of multiple imputation of the outcome, which relies on specific assumptions to produce unbiased estimates. Without this information, it is not possible to evaluate whether the imputation approach was appropriate. (10.1093/ije/dyz032).

Response:

We have added the information regarding the assumption for the multiple imputations for missing data in the Methods section:

Multiple imputations for missing data were conducted under the assumption that data were Missing at Random (MAR). Under MAR, the probability of missingness is related to observed variables (such as age, baseline functional score, and chronic conditions) already included in the models. We employed the fully conditional specification (FCS) method for the imputation of outcomes and covariates. This method imputes each variable with missing values using a regression model conditional on the other variables, iterating through all variables until convergence. It is flexible and easy to implement, allowing variable-wise models and accommodating different types of variables. All covariates were included in the imputation models to create 10 imputed datasets. To improve the validity of the MAR assumption, we included all longitudinal outcomes, baseline covariates, and mortality status in the imputation models, as recommended by Hughes et al. to reduce bias in longitudinal settings.[27]

5. The modelling strategy remains unclear. The authors appear to have both linear versus non-linear relationships and subsequently employed restricted cubic splines. I was not able to understand why the authors used multiple trajectory description. The rationale for combining both methods is not explained, and readers cannot readily follow the analytical procedures that were conducted.

Response:

We have added explanation for this in the Methods section:

Recognizing that while physical function declines with age, stroke may accelerate this deterioration, we hypothesized that these trajectories would involve both linear and non-linear patterns. To test this, we first established overall linear trends using fixed-effect coefficients. We then employed likelihood ratio tests to compare these linear models against non-linear alternatives using restricted cubic splines, allowing us to identify the precise timing of accelerated decline. The optimal number of knots was determined using the Akaike and Bayesian information criteria (Tables S14-S17). We tested models with two to eight knots for the trajectories of individual scores. The annual changes in each score for each time period between knots were analyzed using linear mixed models with marginal means. This dual approach—combining splines with period-specific marginal means—was designed to provide a continuous visual representation of functional change alongside quantifiable, clinician-friendly estimates of 'annual change' during specific windows. Consequently, we were able to confirm the non-linearity of the trajectories while quantifying the rate of decline during critical periods.

Reviewer #3 (Remarks to the Author):

The authors significantly improve the quality of results and the way they are presented. I have only some minor comments to report:

-Revise some expressions and soften the tone of some sentences: e.g. "The observation that ADL and IADL trajectories begin to diverge several years before a stroke... suggests two intervention windows." Evaluate to replace with "two potential intervention windows". Change "there may be a bidirectional relationships" with "there may be a bidirectional relationship".

Response:

We thank the Reviewer for these helpful suggestions and have incorporated them into the revised manuscript.

-The phrase "Several years before and after" appears frequently. Considering replacing some with ""in the peri-stroke period", "around the time of stroke", "in the years surrounding stroke".

Response:

This has been revised accordingly.