nature portfolio

Peer Review File

Brain change trajectories in healthy adults correlate with Alzheimer's related genetic variation and memory decline across life

Corresponding Author: Dr James Roe

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)
The authors present a comprehensive multi-cohort study of associations between genetic (polygenic) risk for Alzheimer's disease (AD), brain changes and cognitive functioning. They report associations between AD genetic risk and accelerated brain volume loss, and between brain change, AD genetic risk and memory decline. The study contains original findings that are of significance to the field of neurodegeneration research, is expertly conducted, and documented with the appropriate level of detail.
In terms of methods I have one main question: the analyses were largely conducted on data from multiple scanners, even across field strengths (1.5T and 3T). How was this considered in the analysis? The authors mention a "scanner covariate indexed study cohort" for one of the analyses, but it is not clear which dataset(s) this refers to, and how the information about scan site/ cohort/ field strength was incorporated in the analysis. Please explain how this was done (or if it was not done for some analyses, please explain why it was not necessary).
Conceptually, I was not convinced by the authors' final conclusion, that "Our results call for a dimensional approach to late onset AD as not being clearly distinct from normal brain ageing." The main finding of their study was an acceleration ("faster than expected") of brain volume loss with genetic AD risk. Without longitudinal data that allow for a separation of the cohort into participants who do and those who do not develop clinical AD it is very difficult - perhaps impossible - to decide the old question whether AD can be described as accelerated ageing (and thus, perhaps, not "clearly distinct from normal brain ageing" although I would be careful to use this wording even then), or whether it amounts to a distinct pathological process.
Several remarks:
1. I would suggest the authors go through their ms. for general understandability – a sentence such as the following is very difficult to parse:
"These data contradict a recent GWAS finding the effect of APOE upon amygdala and hippocampal slopes, with increasing influence of the APOE-indexing SNP (rs429358) with age, disappeared after accounting for disease in a heavily patient-derived sample, suggesting APOE-mediated slope differences were driven by patients."
Another example is "Of note, though PRS-AD effects were not entirely driven by allelic variation in APOE, PRS-ADnoAPOE associations were most evident using the genome-wide significant SNP's/weightings reported by Jansen et al. or Lambert et al., suggesting these SNP sets beyond APOE better capture differences in brain ageing in healthy adults"
Or: "In our case, the principle arbitrary covariate was the age-range to test the association across, and the influence of this arbitrary choice on statistical significance is made clear in Fig. 2, Fig. 4B and Fig. 5, despite accounting for age- and time-related covariates"
2. The authors write: "While these are well-validated and reliable, it is possible measures such as entorhinal cortex may be less reliable" – there is more Freesurfer QC literature, which may address the authors' concerns
3. I would not speak of the "presence of AD risk genes" but of "risk variants"

Reviewer #2

(Remarks to the Author)
This is an interesting manuscript in which the authors explore the association between common allele risk for Alzheimer's

disease (AD; polygenic risk scores based on 4 different GWAS) and brain morphometry and cognitive ability. The novelty of their approach lies in the longitudinal approach, with most participants producing over 2 time points of data, and the calculation of an individual's index of deviation from 'normative' atrophy trajectory. The rationale behind being that AD risk would associate with a quicker aging/atrophy of the brain. The goal of the study is not that clearly spelled out in the introduction, though, even more so when it comes to the cognitive assessment. I think the manuscript would benefit from a more clear and direct presentation of the main goals of this study with regards of both the brain morphometry measures and the cognitive assessment; as well as making clear what the real contribution of these would be in terms of what we know about AD or how we treat it.

With regards of the methods, I wondered how time-point contribution distributed across age. In other words, is it the case that older participants were scanned mostly only 2 times and therefore the calculation of the trajectories is affected by less longitudinal data at older ages? That along the reduction in sample size with age, could explain the increase in variability in older age groups and affect the robustness of the results.

As per the exclusion due to neurodegenerative, neurologic or psychiatric disorders – and more specifically relating to the last – how was that assessed, and did it apply as 'presence' or 'lifetime'? Would also be good to clarify 'medication known to affect the CNS', since this could be quite an extensive list and I presume the authors here refer mostly to medications aimed at treating the above conditions.

Two extreme outliers for the 'chage' data from each one of the replication samples were excluded (supp Fig 8). These were certainly extreme outliers (>5SD) visually detectable in that figure. However, it appears that other participants show extreme – albeit less extreme – values. Did the authors perform any examination of data distribution and used any outlier criteria to clean up the samples (including the sample used for their main analyses)?
Did the authors apply any quality control to imaging data obtained from Freesurfer, or all outputs were taken into the analyses? Freesurfer is susceptible to produce wrong readings for hippocampus, amygdala and entorhinal cortex. These are ROIs of particular interest here.

Most participants from NESDA presented with psychiatric conditions, which would have been expected from that cohort. There is no indication as per how many from BETULA presented with mental conditions, though? Also, considering that for their main analyses the authors excluded participants with mental conditions; would not have been better to apply the same criteria here? otherwise, what could be the effects of this on the results obtained?

Lines 734-735: "we reasoned that PRS' constructed with more relaxed p-value thresholds will be less comparable across the four scores". Can the authors please explain this, I do not see why? In this respect, I would have thought that the authors only retained those SNPs common across all samples and the GWAS studies, but I could not find mention of this. Was this done? It would also be very helpful a table showing the number of SNPs included in each PRS calculated. P-value thresholds in line 736 have gone mixed with references, I presume those are $p<1*10-5$ and $p<0.1$? Considering the method that the authors have used to calculate their PRS, I would rather see a spread of p values, rather than only those 3 very distant thresholds chosen. Considering that their main analyses are on healthy controls, I do not quite follow the rational to use p-thresholds that have shown better to discriminate between patients and controls.

Could the authors add in the methods how to interpret the sign of the 'age relative change' value. That would substantially help the reading. Also, can they add how to interpret 'absolute change'? Does a positive score in 'absolute change' indicate increased in volume? If so, how do the authors interpret the large number of participants whose entorhinal cortex 'grows' in between the ages of 30 and 70? Is this result plausible, or I am interpreting this metric wrongly?

Sorry, I could not make sense of Figure 1d with the text provided in the Figure legend or in the main document. The fact that the relative-age change can take positive or negative values with a different meaning (as I understood this metric) makes the interpretation of its association with PRS AD not that straight forwards. I thought more needs to be said about the interpretation of this result.

I am puzzled by the function representing absolute change across age for cognition. Can the authors please explain this further and how do they interpret the raise in between 50 and 70 years of age?

Overall, I thought that the methods applied were too succinctly explained giving their complexity, which will make the understanding of the results obtained difficult to interpret for most readers. I think the manuscript will benefit from more clarity in this respect, considering the wide audience that Nature Communications aims to reach. In general, this manuscript was hard to read, and would benefit from proof-reading.

"Genetic AD risk is robustly associated with the slope of brain ageing in healthy adults" -> I feel this is an overstatement. It is true that the authors provide some evidence for this, but I would not call that 'robust'. The authors certainly report results convincingly associating AD PRS with hippocampal volume; however, when APOE-e4 carrier status is accounted for, the associations clearly drop, and no interaction agexAD-PRS is found to survive multiple testing in the full-age range analysis. Results, become far less consistent for their analyses of ROIs for Braak stages I and III.

"By specifically isolating within-individual genetic effects on accelerated brain ageing, the present study confirms AD risk genes also influence normal variation in hippocampal change rates in healthy adults." I don't think this conclusion is correct, as far as I understood the authors did not limit their polygenic scores to SNPs within risk genes, or even exomes; so, I don't think any conclusion can be drawn with regards of where this signal is coming from within the genome (may well be driven

by SNPs outside genic regions).

I found very little justification as per why the authors applied other p-thresholds than the genome-wide originally selected throughout the reading of the results and discussion. As mentioned earlier, I do not really understand why to use such stringent p-threshold when calculating a PRS, since the authors are rejecting a lot of valid information; and I did not find in the paper any convincing argument to such unusual strategy. Likewise, there is no clear justification or interpretation for the use of 4 different GWAS results. On a side note, were the participants included in this study, also included in any of those GWAS?

In general, I found the discussion overstated and would suggest downplaying the conclusions of this study.

Reviewer #3

(Remarks to the Author)
This is a very impressive article that asks some fundamental questions about trajectories of structural decline in healthy adults compared to participants with AD from the Alzheimer's Disease Neuroimaging Initiative (ADNI). The authors used machine learning to characterize brain regions that declined most steeply with AD. Then they isolated an adult lifespan sample of normal adults aged 30-90 and assigned each subject a "polygenic risk score (PRS)" based on APOe4 status and other genetic markers of AD. This set the stage to allow them to test, using multiple methods, whether those with higher risk scores showed a pattern of volume loss similar to the AD patients. Using a sliding window that progressively narrowed their sample by age, they determined that an accelerated pattern of change similar to AD patients typified those with higher PRS scores. This is a very impressive article that asks some fundamental questions about trajectories of structural decline in healthy adults compared to participants with AD from the Alzheimer's Disease Neuroimaging Initiative (ADNI). The authors used machine learning to characterize brain regions that declined most steeply with AD. Then they isolated an adult lifespan sample of normal adults and aged 30-90 and assigned each subject a "polygenic risk score (PRS) based APOe4 status and other genetic markers of AD. This set the stage to allow them to test, using multiple methods, whether those with higher risk scores showed a pattern similar to the AD patients and an accelerated rate of decline. Using a sliding window that progressively narrowed their sample by age, they determined that an accelerated pattern of change similar to AD patients typified those with higher PRS scores. The analyses are impeccable, thorough, and I would actually describe them as elegant.
I think this study unquestionably validates the use of PRS scores as measures of risk. But that is not really the aim of the study. The introduction focuses on whether normal aging and AD follow similar trajectories but at different rates. It seems to me that the theoretical and empirical work that would be most informative would be to use the powerful and creative methods described to determine the youngest age at which a predictive pattern of polygenic factors could validates the use of polygenic markers in assessing risk, but this be found. As it stands now, by selectively decreasing the sample from youngest to oldest, the authors show that the pattern that is characteristic of older adults is similar to AD patients.

The sliding window approach and accompanying methods seem ideal for examining this issue and could readily be adapted. The findings would be hugely interesting but require a true lifespan sample and a much larger n. Critically, it is a truly unfortunate weakness that the present study, overall has a small number of subjects for a lifespan study and the number of 40-50 year olds is sparse—it appears to comprise only a fraction of the lifespan sample. In general, middle aged individuals are severely under sampled.

I have been considering what pattern of findings might have occurred that would be unexpected. To the extent that the PRS is valid, the result seem preordained. No one comes down with AD acutely and given that a large number of older adults would likely have incipient AD, it doesn't seem that surprising that they have volume loss and decline rates that mimic AD. It is important to keep in mind that adults who have AD or will get it are included in the normal sample. I believe that a study of low amyloid, low PRS adults is necessary to understand what typifies the aging process and what is shared by normal aging and AD. Such a study would likely lead to a large and interesting debate about what constitutes normal aging.

A notable strength of the paper is the combining of different brain structures and examining whether the findings continue to be validated when multiple combinations of structures are tested. Similarly, the finding of more accelerated memory decline coupling with vulnerable, high PRS subjects is another strength. I find this to be an outstanding, landmark study that besides being very informative about AD and some aspects of its etiology, also provides an important template for methods of combining multiple methods and a diversity of analytic techniques to answer some broad questions. of machine language techniques further strengthens the paper.

My only reservation is that I think that the conclusions should not be overgeneralized to lifespan aqing, given the limited sample of middle-aged adults and I think this should be explicitly noted as a weakness. The authors also correctly note the limitations of not having measures of amyloid available.

Despite a few concerns. I find this to be an outstanding, landmark study that besides being very informative about AD and some aspects of its etiology, also provides an important template for methods of combining multiple data collection techniques with a diversity of analytic techniques to answer some broad questions. There is unusually good evidence´for replicability ranging from multiple regression to machine learning. The study blends a cognitive aging approach to a more traditional medical sciences approach. I believe we will only solve the AD problem if these groups join forces, and this collaboration clearly shows how to do it. I do recommend that the uneven age sampling, particularly middle-aged, should be explicitly noted as a limitation in making general comments about aging. The authors also correctly note the limitations of not having measures of amyloid. This study gave me great hope and optimism about the study of AD and aging. Regardless of

the editorial decision, the authors are to be congratulated
For their forward thinking and excellent model for achieving scientific advances.

Reviewer #4

(Remarks to the Author)
This research aims to investigate the effect of genetic Alzheimer's disease (AD) risk on brain changes during healthy aging. Motivated by the qualitative similarities between healthy brain aging and AD, the authors hypothesize that a higher genetic risk for AD may contribute to faster atrophy rates in healthy adults. Their literature review indicates that previous studies have predominantly relied on cross-sectional designs, resulting in mixed results. Additionally, the reviewed longitudinal studies failed to account for individual differences in brain features. To address these limitations, the authors propose an approach to compute individualized estimates of brain atrophy relative to age. First, they examine genetic AD risk in relation to early Braak regions, known to be affected in both healthy aging and AD. Then, using a data-driven method, they identify additional brain regions exhibiting accelerated atrophy in AD and analyze their associations with genetic AD risk. Finally, this research explores genetic AD risk in relation to individualized estimates of memory change over the course of aging in healthy adults.

Key results:
In a cohort of healthy adults, polygenic risk scores for AD (PRS-AD) are associated with reductions in key brain features affected in AD, specifically hippocampal volume, and a data-driven multivariate marker of brain change. Interestingly, these associations are largely, though not exclusively, driven by APOE. Finally, among healthy adults who are at high genetic risk for AD, those with higher rates of brain change exhibit greater memory decline relative to those with less brain change.

Validity
The claimed results of the study are:
1. PRS-AD are associated with greater age-relative change in early Braak regions (the hippocampus, amygdala, and entorhinal cortex). These extend beyond the contribution of APOE.
2. a) PRS-AD is significantly associated with change in a multivariate marker of AD-accelerated brain features in healthy adults.
b) Some of the PRS-AD associations with brain change in healthy adults are replicated in an independent sample.
3. Among individuals with high PRS-AD, those who are high on the multivariate marker of brain change present greater reductions in memory over the adult lifespan relative to those with lower brain change.

Claim 1: This result is most convincing in the left hippocampus, where multiple PRS-AD scores are associated with age-relative change and absolute brain change for most of the age-ranges. 86% of these survive FDR-correction and 33% remain significant when APOE is taken out of the PRS-AD. The associations with the right hippocampus and right amygdala are largely driven by APOE. The associations with the entorhinal cortex and the cortical ROIs of Braak Stage III are much weaker/non-existent after FDR correction.

Claim 2a: 14 of the 36 tested associations relating the multivariate marker of brain change to PRS-AD were significant after FDR-correction, but the results are largely driven by APOE (one association remains significant without APOE, in the age group 65+) It is interesting that the sliding-window PCA yielded similar results, ensuring that no specific brain regions are driving the multivariate effect. It was thoughtful to run the analyses with and without APOE, as well as with a sliding window PCA as this facilitates interpretation.

Claim 2b: The association with the left hippocampus is replicated, and age ~50 is again found to be the time when brain changes increase. However, the PRS-AD associations with the amygdala, and the multivariate marker of change are not replicated. The latter is a key finding in the study, and the lack of replication is an important limitation. The authors mention the replication sample not having as many follow-ups, but elaborating on other explanations would be helpful.

Claim 3: This finding is interesting, especially because it is not driven by APOE e4 carriership. It suggests that individuals with both a higher genetic risk for AD, and a higher index of brain change are more likely to exhibit greater memory decline over the course of aging. The analyses, however, also suggest the brain change marker has more predictive power than PRS-AD, thus undermining the relevance of looking at genetic AD risk.

Significance
This research highlights the parallels between healthy aging and AD and suggests that neurodegeneration occurs along a continuum. This holds clinical significance, especially given the growing interest in combination therapies that target the biology of aging in AD. The finding in relation to memory decline further emphasizes the translational significance of the paper. Finally, the findings add to the lifespan neuroscience literature while addressing methodological gaps.

Data and methodology
The data samples appear to have been carefully selected and preprocessed.

Analytical approach
The Generalized Additive Mixed Model is used to compute individualized estimates of change in brain features over aging. The specified models are reasonable, and this approach allows for individualized estimates of brain change over time. The XGBoost binary classifier is trained on brain change features to differentiate those with an AD diagnosis at the final timepoint from those with normal cognition across all time points. The model is validated externally on AIBL. The approach is reasonable, and the chosen model seems to have traded specificity for sensitivity.

Suggested improvements
1. Figure 2, 4, 5: it is difficult to interpret what it means for one PRS-AD score to remain significantly associated with brain change, while another does not after FDR correction and/or removal of APOE. A brief explanation of the four summary scores, and how they differ from one another would be helpful in the introduction.
2. Authors could explore additional modeling architectures in addition to XGBoost. This would help ensure that the findings are consistent and relatively agnostic to model frameworks.
3. Additional performance metrics should be reported, such as area under the precision-recall curve. The mean results over the 500 iterations of the 10-fold cross-validation should be reported as well.
4. The lack of association with the entorhinal cortex is surprising. All brain features were measured via FreeSurfer, which, according to the authors, limit accurate measurements of the entorhinal cortex. They mention manual tracing as a potential solution, but this may not be feasible in a large multi-cohort dataset. Alternative solutions could include other processing toolboxes (such as CAT12) and atlases (such as Hammers_mith) or extracting radiomic features, and then comparing the results with the FreeSurfer measures.
5. The authors should discuss the potential effects of scanner type, field strength and sequence parameters.
6. The sample's breakdown by race and ethnicity is not reported; if the sample is homogenous, it should be mentioned as a limitation in the discussion.
7. Regarding the replication analyses: since APOE seems to be the primary factor driving the associations between genetic AD risk and brain change, it is possible these associations are not significant in the replication sample if it happens to have fewer participants who are APOE-e4 carriers. It could be helpful to look at the distribution of the APOE genotype in the replication sample to help explain the lack of replication.
8. The clinical implications of the study could be clarified.
a. Do the results suggest that healthy individuals with higher PRS-AD and higher age-relative brain changes may develop mild cognitive impairment or dementia in the future? Is it possible these clinically healthy individuals may already be on the biological AD trajectory? ADNI and AIBL have biomarker data, which could be used to gain insight into the mechanisms underlying the observed associations.
b. Is it only fair to say that individuals with greater age-relative changes in brain volumes are likely to have a higher PRS-AD, or does this association mean these individuals are likely to develop dementia? The authors touch on this limitation at the end of the discussion, but it could be helpful to clarify the implications of the study despite this limitation.
c. The authors mention that they disregarded AD heterogeneity as this was reasonable for their purpose of identifying brain features that tend to change faster in AD. However, it is possible that the "age-relative change" metric is capturing changes driven by different subtypes rather than by individual differences. It could be helpful to run additional analyses accounting for the major AD subtypes.

Reviewer #5

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports as part of the Nature Communications initiative to facilitate training in peer review and appropriate recognition for co-reviewers

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
All my comments have been addressed in a satisfactory manner.

Reviewer #2

(Remarks to the Author)
Having had the opportunity to read the revised version of this manuscript, along with the author's detailed responses to my comments, I am happy that most my questions have been resolved. As I understand the main contribution of this study is showing that AD common genetic risk associates with an accelerated general brain ageing, more prominent, though, in the hippocampus. This association not being only driven by the APOEe4, despite this being a strong contributor to it. The methods used, the results presented and the conclusions regarding this appear clearer to me now and I have no major concerns about it. I would only suggest rephrasing the author's second hypothesis that, as written, does not read well to me (lines 148-149: "and that variation in AD risk factors would be detectable via brain change rates relative to one's age in AD-sensitive regions in healthy adults") -> which AD risk factors are the authors referring to? And what 'being detectable mean'? I think here they were to suggest that AD PRS will be predictive of a faster decline than expected by age in AD sensitive structures like hippocampus; but not sure it is this by the reading of their statement.

The section of the manuscript that I still struggle a bit with, is the analyses involving cognitive function. Admittedly, I do not fully understand the method that the authors applied, but my concern goes beyond that, I don't understand it's purpose and what is its real contribution to the overall study. First, I do not see the link between this section and the main results/conclusion; in fact, the authors do not devote much to it either: it is only marginally introduced/justified at the end of the introduction, and it is only discussed briefly with no much emphasis on what brings to the table. Initially I thought this section was aimed at answering the question: 'is it the accelerated brain decline linked to AD genetic risk key in explaining AD cognitive decline, or the speed of decline is secondary to the absolute brain volume?' In other words, does it matter cognitively whether you lose volume faster than expected for your age, or what matter is whether you lose up to certain total amount of grey matter. However, the analyses performed do not allow to test the above, but their four-group breaking of the sample lets to the conclusion that brain changes are potentially the key element predicting cognitive performance (line 505-506) than genetic risk per se; which should not come as a surprise, since one could predict genetic risk to impact cognitive phenotypes via its potential effects on brain?? Admittedly, it might be me, but I can't really make much sense of what the authors intend to explore here and what the novelty of their results is. Perhaps this should have been spelled more clearly? Allegedly, though, this is a small part of their work that should not take away the merits of the main analyses/results.

Reviewer #3

(Remarks to the Author)
I FOUND THIS MANUSCRIPT TO BE MUCH IMPOVED OVER THE ORIGINAL WITH GREATER CLARITY, PARTICULARLY IN THE INTRODUCTION AND DISCUSSION. I WAS IMPRESS2ED AT HOW THE AUTIHORS RESPONDED COMPLTELY, YET CONCISELY, TO THE MANY PAGES OF COMMENTS THEY RECEIVED. I THINK THIW WAS A JOB WELL DONE OIN EVERY IDIM ENSION.

Reviewer #6

(Remarks to the Author)
Authors have responded to all of the recommendations and have clarified the raised concerns.

Version 2:

Reviewer comments:

Reviewer #2

(Remarks to the Author)
The authors have satisfactorily replied to all my queries. I believe the authors have now better justified the final analyses and made clearer the contribution of these to the overall paper. Some of the points raised in their reply could be debated and I am sure will stimulate further discussion and research ideas, which I see as a possitive. I believe this manuscript represents an important contribution to the field.

We thank the reviewers for their insightful comments that have helped strengthen our paper. Our point-by-point response is below (in green). Changes are highlighted in red and in the revised manuscript (note we made many changes also to comply with formatting guidelines).

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

The authors present a comprehensive multi-cohort study of associations between genetic (polygenic) risk for Alzheimer's disease (AD), brain changes and cognitive functioning. They report associations between AD genetic risk and accelerated brain volume loss, and between brain change, AD genetic risk and memory decline. The study contains original findings that are of significance to the field of neurodegeneration research, is expertly conducted, and documented with the appropriate level of detail.

1.1. In terms of methods I have one main question: the analyses were largely conducted on data from multiple scanners, even across field strengths (1.5T and 3T). How was this considered in the analysis? The authors mention a "scanner covariate indexed study cohort" for one of the analyses, but it is not clear which dataset(s) this refers to, and how the information about scan site/ cohort/ field strength was incorporated in the analysis. Please explain how this was done (or if it was not done for some analyses, please explain why it was not necessary).

Thank you. The Reviewer is correct that the discovery cohort originated from 3 different scanners. We apologize it was not clear how this was dealt with. In this cohort, we corrected for scanner in the GAMM that estimated subject-specific brain change. We have made this clearer (Methods):

"fitting a nonlinear term for age ~~(corrected for sex, scanner, and intracranial volume)~~, with covariates added for sex, scanner, and intracranial volume"

The part the reviewer refers to concerns the replication analysis. That is, we wrote that we ran comparable GAMMs in the replication sample. But that since each of the two studies comprising the replication sample originated from a single scanner, the scanner covariate now indexed study cohort. We made this more explicit:

"We first ran a GAMM separately in each of the replication cohorts, revealing a strong outlier for each in the hippocampal change data (-7.4SD in BETULA; +5.5SD in NESDA; see SI Fig. 8). Then, we collated the data from the two cohorts, ran a GAMM comparable to the main analysis estimating the random slopes, and excluded these two outliers. Note that since each of the two cohorts originated from a single scanner, the scanner covariate indexed study cohort."

For ADNI data, we did not allow subjects to change field strengths across timepoints. We did this to ensure the validity of subject-specific change estimates in these legacy data, and were more at liberty to do so in ADNI due to the number of available longitudinal scans. We added the following clarifications in Methods and Results:

"After grouping, for subjects where scanner field strength changed over time (from 1.5T to 3T), we used only the observations from the scanner with the most timepoints (or where equal used the 3T scans). This was to ensure the validity of change estimates in the multiscanner, multisite ADNI data[1]"

"Then, in 364 features we modelled a GAMM of age (irrespective of group), and entered the individual-specific slopes into ML binary classification (Fig. 3B; model covariates: sex, scanner field strength, and ICV; note that the ADNI sample did not change field strength over time; Methods)."

1.2. Conceptually, I was not convinced by the authors' final conclusion, that "Our results call for a dimensional approach to late onset AD as not being clearly distinct from normal brain ageing." The main finding of their study was an acceleration ("faster than expected") of brain volume loss with genetic AD risk. Without longitudinal data that allow for a separation of the cohort into participants who do and those who do not develop clinical AD it is very difficult - perhaps impossible - to decide the old question whether AD can be described as accelerated ageing (and thus, perhaps, not "clearly distinct from normal brain ageing" although I would be

careful to use this wording even then), or whether it amounts to a distinct pathological process.

We agree we should have been more careful here. We have changed the Conclusion, and expanded on this point in the Discussion:

"In conclusion, brain ageing trajectories in healthy adults are accelerated by the presence of AD risk variants, in many brain features, and in some cases beyond *APOE*. We show that brain features most susceptible to faster deterioration in AD are on a trajectory of accelerated change from age ~50 in healthy individuals, and that models trained on AD patients can be applied to adult lifespan data and the prediction relates to genetic AD risk. Finally, genetically at-risk individuals with more brain change showed more adult lifespan memory decline, compared to genetically at-risk individuals with less brain change. Thus, tracking brain change in AD-sensitive regions enhanced the value of knowing a person's genetic risk, and atrophy predicted memory decline more than the genetics. Our findings show that brain ageing slopes in healthy adults correlate with AD-related genetic variation and memory decline through adulthood, and that neurodegeneration occurs along a continuum from normal ageing to AD – supporting a dimensional, lifespan view of AD."

"Fourth, we used only structural measures. While these are sensitive to detecting subtle changes in brain structure that ultimately reflect a continuous, lifelong process of change, other biomarkers are necessary to refine detection of AD-risk in healthy adult samples. Finally, we do not know which individuals here will be diagnosed with AD later in life, or have other AD biomarkers suggesting a biological trajectory to AD[2]. While our analyses suggest one could assign differential transition probabilities to healthy individuals, only time and follow-up data will tell."

Several remarks:
1.3. I would suggest the authors go through their ms. for general understandability – a sentence such as the following is very difficult to parse:
"These data contradict a recent GWAS finding the effect of APOE upon amygdala and hippocampal slopes, with increasing influence of the APOE-indexing SNP (rs429358) with age, disappeared after accounting for disease in a heavily patient-derived sample, suggesting APOE-mediated slope differences were driven by patients."
Another example is "Of note, though PRS-AD effects were not entirely driven by allelic variation in APOE, PRS-ADnoAPOE associations were most evident using the genome-wide significant SNP's/weightings reported by Jansen et al. or Lambert et al., suggesting these SNP sets beyond APOE better capture differences in brain ageing in healthy adults"
Or: "In our case, the principle arbitrary covariate was the age-range to test the association across, and the influence of this arbitrary choice on statistical significance is made clear in Fig. 2, Fig. 4B and Fig. 5, despite accounting for age- and time-related covariates"

We have made extensive amendments to increase readability (see highlights in ms). The examples referred to:

"These data contradict a recent GWAS, which found the effect of *APOE* upon amygdala and hippocampal change in ageing disappeared after accounting for disease[3]."

"Most associations after excluding *APOE* were with scores derived from the genome-wide significant SNP's/weightings reported by Jansen et al.[4], possibly suggesting these better capture differences in brain ageing (though PRS-AD^{noAPOE} effects were also evident with scores from two other GWAS[5,6])"

"In our case, the main arbitrary covariate was the age-range to test the association across. Despite always accounting for age- and time-related covariates, the influence of this choice upon statistical significance is clear in Fig. 2, Fig. 4B and Fig. 5."

1.4. The authors write: "While these are well-validated and reliable, it is possible measures such as entorhinal cortex may be less reliable" – there is more Freesurfer QC literature, which may address the authors' concerns

We conducted a QC analysis of the cortical change estimates based on this and a later comment. Please see SI Figure 3 and our response to comment 2.10. While we agree FS in general produces reliable estimates, the evidence points to poorer entorhinal reliability. We added the following Discussion about this:

Third, we relied on FreeSurfer-derived measures. While these are well-validated and reliable[7–9], some measures may be less so[9]. Indeed, that we observed no PRS-AD associations with left entorhinal change was surprising. When we quantified the proportion of individuals estimated to show positive absolute change cortex-wide (i.e., "growth"), entorhinal measures were clear outliers, with ~16% estimated to be growing (Supplementary Fig. 2). The median across cortical regions was 0.2%. This

may reflect poorer reliability of entorhinal measures, as suggested by others[8,10,11]. Possibly, manual entorhinal tracing may have led to different results[9]

1.5. I would not speak of the "presence of AD risk genes" but of "risk variants"

We agree this is more accurate, and have changed all instances. Thank you for your comments.

Reviewer #2 (Remarks to the Author):

This is an interesting manuscript in which the authors explore the association between common allele risk for Alzheimer's disease (AD; polygenic risk scores based on 4 different GWAS) and brain morphometry and cognitive ability. The novelty of their approach lies in the longitudinal approach, with most participants producing over 2 time points of data, and the calculation of an individual's index of deviation from 'normative' atrophy trajectory. The rationale behind being that AD risk would associate with a quicker aging/atrophy of the brain.

2.1. The goal of the study is not that clearly spelled out in the introduction, though, even more so when it comes to the cognitive assessment. I think the manuscript would benefit from a more clear and direct presentation of the main goals of this study with regards of both the brain morphometry measures and the cognitive assessment; as well as making clear what the real contribution of these would be in terms of what we know about AD or how we treat it.

To make the aims of the study clearer, we added the following in the Introduction:

"Here, we aimed to the test the hypothesis that brain change trajectories in healthy adults correlate with AD-related genetic variation and memory decline through adult life. We hypothesized that neurodegeneration in ageing and AD is linked on a continuum, and that variations in AD risk factors would be detectable via brain change rates relative to one's age in AD-sensitive regions in heathy adults. We further hypothesized that individuals with both higher genetic risk and faster atrophy for their age would exhibit more memory decline through adulthood."

We also added text in the Discussion to discuss the possible translational implications:

"Moreover, group differences in memory decline were more driven by brain change differences than by genetic differences, as they persisted when controlling for PRS-AD and *APOE-ε4* but not atrophy. Thus, brain change may be crucial for detecting comparatively at-risk individuals in adult lifespan data. Further, memory decline differences were protracted through adulthood, as they were evident in different age subsets, including comparatively younger adults (e.g., within the age-range 30-65; Supplementary Fig. 17). This indicates neurodegeneration in AD-sensitive regions tracks with memory decline differences that are detectable through adulthood. It also emphasizes that memory decline is a gradual phenomenon not confined to old age. This perspective may be obscured in studies that use clinical tests that do not capture subtle cognitive variations, and estimates of decline relative to a group rather than one's earlier capacity. Our findings extend previous studies finding PRS-AD[43,44,46] or *APOE-ε4*[45] relates to memory decline across adult life, and possibly shed light on why associations are often weak[43–46] or absent[49]. Whether these brain trajectory differences are relevant for later AD outcomes will require follow-up and biomarker assessment, but our results show these neurodegenerative changes are not benign. They also underscore the need for follow-up data over extended age-spans for prediction or prevention of AD, and suggest a continuous view on lifespan brain health may aid understanding of AD[78,79]. Multivariate atrophy measures may help assess AD risk and improve selection into clinical trials."

"Sixth, we used only structural measures. While these are sensitive to detecting subtle changes in brain structure that ultimately reflect a continuous, lifelong process of change, other biomarkers are necessary to refine detection of AD-risk in healthy adult samples. Finally, we do not know which individuals here will be diagnosed with AD later in life, or have other AD biomarkers suggesting a biological trajectory to AD[2]. While our analyses suggest one could assign differential transition probabilities to healthy individuals, only time and follow-up data will tell."

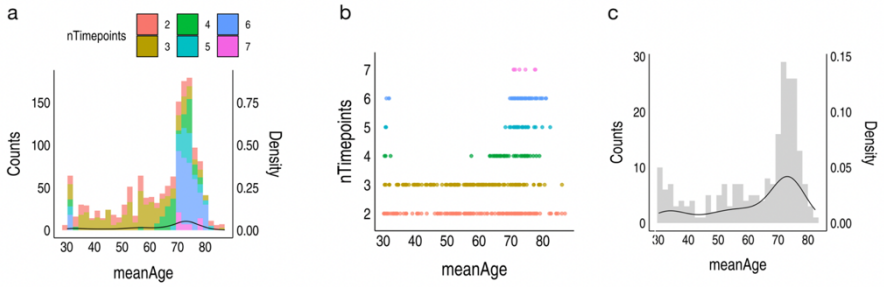
2.2. With regards of the methods, I wondered how time-point contribution distributed across age. In other words, is it the case that older participants were scanned mostly only 2 times and therefore the calculation of the trajectories is affected by less longitudinal data at older ages? That along the reduction in sample size with age, could explain the increase in variability in older age groups and affect the robustness of the results.

This is important to clarify. It is the opposite; we have more longitudinal data at older ages, rendering older trajectories possibly more accurate. We have made this and other considerations around data

characteristics and the trajectory estimation procedure explicit. We also added SI Figure 6 showing the timepoint distribution. The increase in variability with older age is a known phenomenon we now also refer to [17]:

"For each of our *a priori* ROIs, we used the random slopes as response variable in linear models with a PRS-AD predictor and the following covariates: mean age (across timepoints), sex, the first 10 genetic PCs (GAFs), number of timepoints, and the interval between first and last timepoint. There was a tendency for older adults to have more timepoints (Supplementary Fig. 6). Hence, the latter two covariates helped ensure the effects were not driven by the uneven distribution of timepoints across age."

"There are also limitations. First, characteristics of lifespan data will affect the mixed-model change estimation. Our study had more timepoints in the older age-ranges (Supplementary Fig. 6), likely resulting in more accurate estimates in older adults. Hence, alongside mean age, we corrected for timepoints and the interval between first and last visit to ensure the results were not driven by residual age-related variation. Similarly, normalizing change estimates by age does not reduce variability along the age variable. Because more variability in older age is a known phenomenon[17], adjusting for mean age further ensured the results were not driven by higher dispersion in older adults. Conversely, estimates in younger, less-sampled age-ranges may be biased by the magnitude of change in older adults. This may help explain why the memory slopes of also younger adults were estimated as negative. Further, selection bias and attrition vary by age-group, which alongside data density differences may explain why adults in their 60's were estimated with less negative memory slopes compared to middle age (Fig. 6B). Caution is thus advised around interpreting change estimates in terms of their absolute values, hence we refer to "estimated change"."



SI FIGURE 6
Age and timepoint distributions. a Age and **b** timepoint distribution of the LCBC discovery cohort used to estimate individual-specific brain change (obs = 1430; N = 420). **c** Age distribution of the genetic observations in this sample (N = 229).

2.3. As per the exclusion due to neurodegenerative, neurologic or psychiatric disorders – and more specifically relating to the last – how was that assessed, and did it apply as 'presence' or 'lifetime'? Would also be good to clarify 'medication known to affect the CNS', since this could be quite an extensive list and I presume the authors here refer mostly to medications aimed at treating the above conditions.

The decision to exclude due to medication use was taken on a case-by-case basis. We have clarified that at baseline, participants were excluded based on lifetime presence of psychiatric disorders or use of medication affecting the CNS, including common psychiatric medicines, with some examples of medication:

"The following exclusion criteria were applied across LCBC studies: evidence of neurodegenerative or neurologic disorders, conditions or injuries known to affect central nervous system (CNS) function (e.g., hypothyroidism, stroke, serious head injury), and MRI contraindications as assessed by a clinician. At baseline, participants were thoroughly screened for evidence of cognitive deficits, and excluded based on lifetime presence of psychiatric disorders and/or use of medication known to affect the CNS (e.g., benzodiazepines, antidepressants or other central nervous agents). Generally, the LCBC sample is comprised of cognitively high-performing individuals (see summary of cognitive scores in SI Table 2; SI Figure 18)."

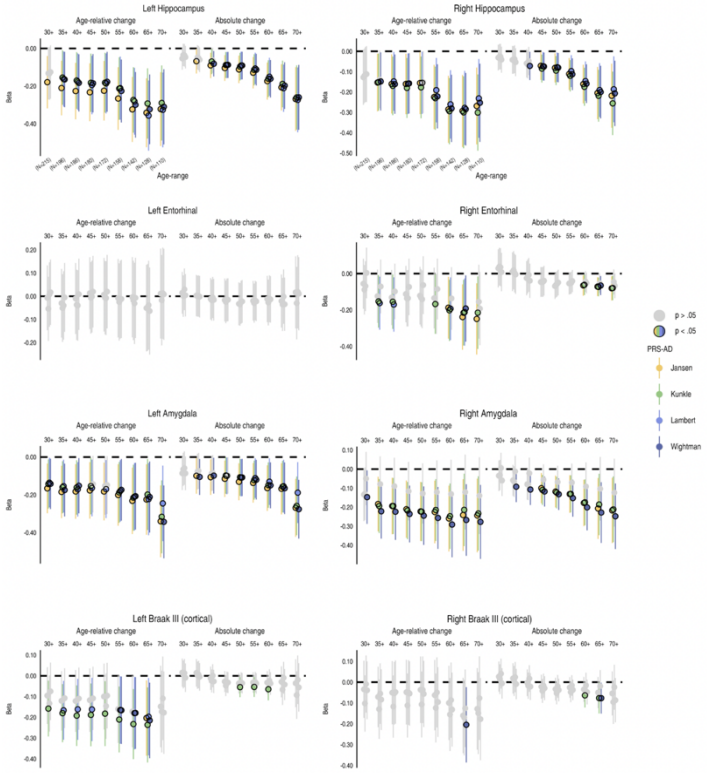
2.4. Two extreme outliers for the 'chage' data from each one of the replication samples were excluded (supp Fig 8). These were certainly extreme outliers (>5SD) visually detectable in that figure. However, it appears that other participants show extreme – albeit less extreme – values. Did the authors perform any examination of data distribution and used any outlier criteria to clean up the samples (including the sample used for their main analyses)? Did the authors apply any quality control to imaging data obtained from Freesurfer, or all outputs were taken into the analyses? Freesurfer is susceptible to produce wrong readings for hippocampus, amygdala and entorhinal cortex. These are ROIs of particular interest here.

MRI images were checked at collection. We did not apply quality control criteria to the brain measures across timepoints, because the statistical methods to estimate change should be robust to outlying datapoints. That is, for subject-specific trajectories with e.g., one outlying datapoint, the mixed-effects estimation will shrink the values of less-probable slopes closer to the mean of the distribution. We added the following to explain this. We also note we typically see that subjects whose measures might be considered an outlier at a single timepoint are often consistently outlying across time (see SI Fig 1).

"The mixed-model change estimation is equivalent to estimating factor scores, and psychometrically superior to more manual calculations of change[18,19]. It should also be less influenced by outliers due to the shrinkage effect. This limits the influence of extreme data points by estimating random effects from a probability distribution, the parameters of which are derived from the data. As more longitudinal measures are incorporated, the distribution becomes more robust, reducing the influence of extreme slopes and pulling them closer to the mean[18,19]. This is exemplified in Supplementary Fig. 13, wherein we show that PRS-AD-change associations in the same individuals in the BETULA study improved when their slopes were estimated together with NESDA data, compared to using BETULA data alone."

Yet the reviewer also refers to outliers in the random effects, which we agree are important to check. We removed all subjects with any slope >3SD on any of the 8 apriori ROIs, and repeated the ROI analysis for the discovery and replication cohorts. This resulted in a max genetic sample of N = 215 (discovery). Across the 576 tests in the ROI-based analysis, the PRS-AD beta coefficients correlated at R=.95 with our reported results. In the replication sample (max genetic sample = 282), the beta coefficients across conducted tests correlated at R=.98. We are thus confident our results are not driven by outliers, and added this as a Sensitivity analysis along with SI Figure 3:

"Importantly, the reported PRS-AD results were not driven by outlying observations (SI Figure 3)."



SI FIGURE 3
Sensitivity analysis: quality control. Univariate associations after discarding data from subjects with observations >3SD in any of the 8 a priori ROIs (maximum genetic N = 215). The set of 576 PRS-AD beta estimates correlated at .95 with the results reported in the main paper.

2.5. Most participants from NESDA presented with psychiatric conditions, which would have been expected from that cohort. There is no indication as per how many from BETULA presented with mental conditions, though? Also, considering that for their main analyses the authors excluded participants with mental conditions; would not have been better to apply the

same criteria here? otherwise, what could be the effects of this on the results obtained?

We added the following in Methods:

"Because it is a population-based sample, BETULA employs no screening/inclusion criteria for mental disorders."

We agree the lack of exclusion for disorders in the replication sample is a limitation. Since the analyses required samples with at least 3 timepoints, we included all NESDA participants, not only their controls, to increase power to statistically estimate change. While we agree one would like to be able to exclude based on disorders also in the replication, the analysis requirements did not provide scope for this. We added the following limitation:

"Fourth, while the discovery sample screened out participants with mental disorders, the replication sample included individuals with disorders (Methods). This decision aimed to increase power to estimate change in the less-powered replication sample, but potentially influenced the results"

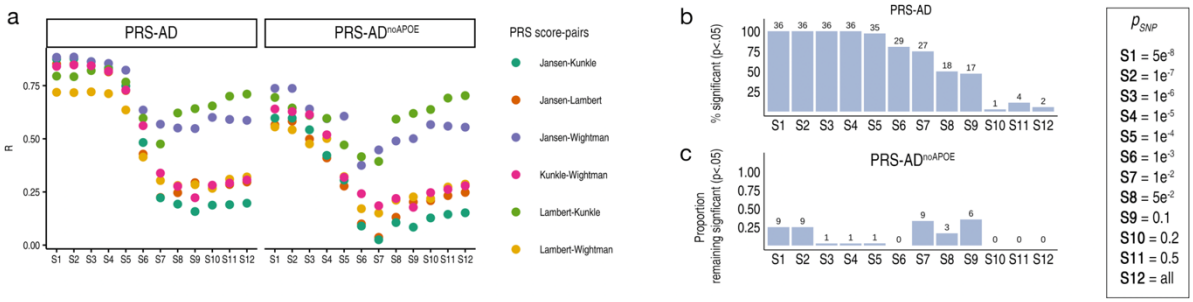
2.6. Lines 734-735: "we reasoned that PRS' constructed with more relaxed p-value thresholds will be less comparable across the four scores". Can the authors please explain this, I do not see why? In this respect, I would have thought that the authors only retained those SNPs common across all samples and the GWAS studies, but I could not find mention of this. Was this done?

PRS' constructed with more relaxed inclusion thresholds are likely to contain different sets of genetic variants. In the ideal situation, with consistent results for the SNP-level effect sizes of a polygenic trait captured across studies, we agree it should not matter. However, the consistency of effect estimates across studies is unknown, and smaller studies will also have larger error in their estimates. Furthermore, a recent critical review of AD GWAS found that despite finding more variants, the more recent studies may explain less heritability, and the increasing sample size may have been bolstered more by including greater numbers of younger controls, by-proxy cases, or cases without clinical screening (Escott-Price & Hardy, 2022). Thus, it is of importance to assess the utility of PRS-AD scores derived from different studies. For these reasons, we opt not to conduct the analyses suggested (i.e., retaining common SNPs). Rather, we added a Sensitivity analysis that empirically supports our claim (SI Figure 4) and text in Methods and Discussion:

Escott-Price, V. and Hardy, J., 2022. Genome-wide association studies for Alzheimer's disease: bigger is not always better. Brain Communications, 4(3)

"We also reasoned PRS' constructed with more relaxed p-value thresholds are more likely to contain different sets of genetic variants, and hence would be less comparable across the four scores. As an exploratory analysis, we tested eleven other p-value thresholds. This confirmed the four scores became less correlated at more liberal thresholds (see Results). Furthermore, adding more SNPs was detrimental to finding PRS-AD associations with hippocampal change in healthy adults (Supplementary Fig. 4)."

"… post-hoc analysis supported our choice of a genome-wide significant threshold for constructing PRS-AD scores; at more liberal SNP inclusion thresholds, the four scores became less comparable, dropping from a median correlation between scores of R = .73 at our chosen threshold, to R = .29 at the most liberal threshold. Moreover, the data suggested including more SNP's was not beneficial, but may be detrimental to finding PRS-AD effects on brain change in healthy adults, at least in hippocampus (Supplementary Figure 4)."



SI Figure 4
PRS-AD correlations and associations computed using 12 alternative p-value thresholds. **a** The correlation between PRS-AD scores dropped when including more SNPs in the PRS, rendering the results between scores less comparable at more liberal thresholds. **b** The percentage count (y-axis) of significant PRS-AD associations with age-relative change in left hippocampus for different PRS computation thresholds (p < .05 [uncorrected]; 9 age-subsets x 4 scores). **c** Where the association was

significant with PRS-AD ($p < .05$ [uncorrected]) we tested whether it remained significant with PRS-AD^{noAPOE} ($p < .05$ [uncorrected]). The plot illustrates the proportion of PRS-AD associations in b that remained significant with PRS-AD^{noAPOE} and their counts.

2.7. It would also be very helpful a table showing the number of SNPs included in each PRS calculated.

We included this as SI Table 2:

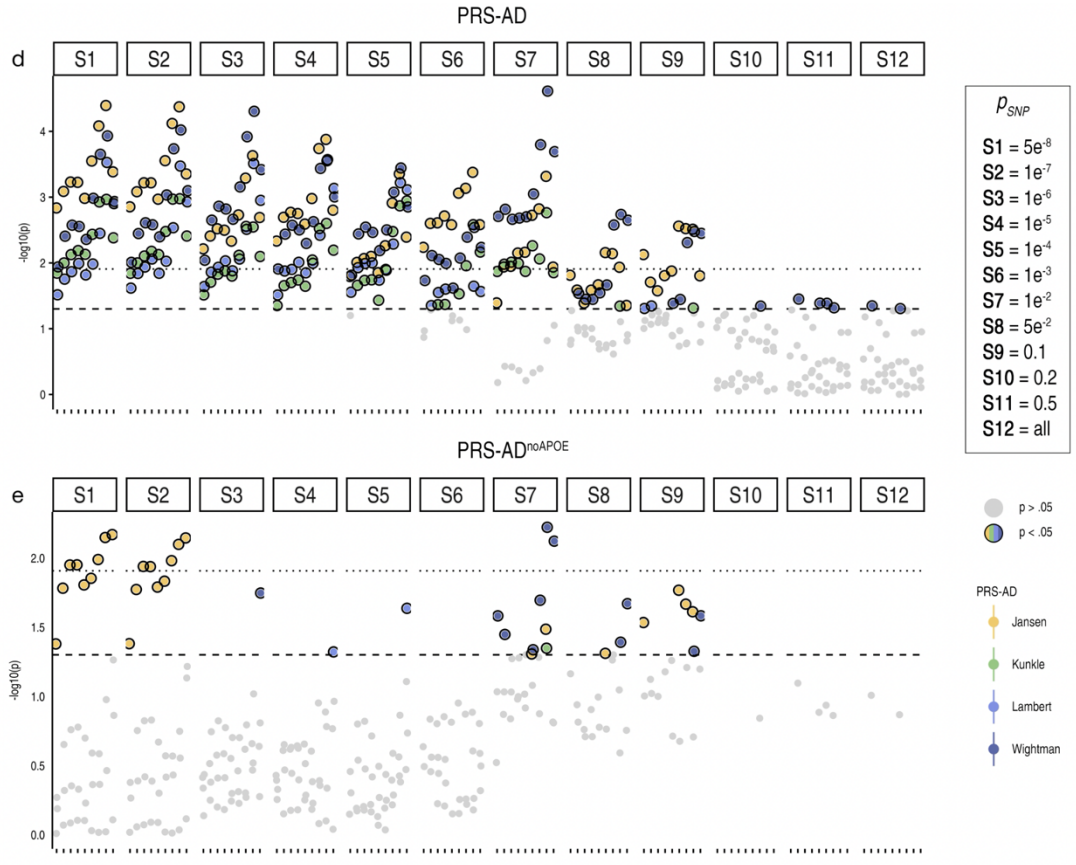
	nSNPs	
Score	PRS-AD	PRS-AD^{noAPOE}
Jansen	59	27
Kunkle	42	16
Lambert	33	13
Wightman	81	42

SI TABLE 2
The number of SNPs used to construct each PRS score.

2.8. P-value thresholds in line 736 have gone mixed with references, I presume those are p<1*10-5 and p<0.1 ? Considering the method that the authors have used to calculate their PRS, I would rather see a spread of p values, rather than only those 3 very distant thresholds chosen. Considering that their main analyses are on healthy controls, I do not quite follow the rational to use p-thresholds that have shown better to discriminate between patients and controls.

We removed the line referred to and followed the suggestion of showing the spread across p-value thresholds. This too is in SI Figure 4 (ctd. below), which also visualizes the associations summarized in the above bar chart. Regarding the rational, we reasoned that if the best threshold for detecting the phenotype in question is X, then our aim to detect genetic variation potentially associated with that phenotype in healthy adults would also benefit from choosing X. After all, the patients were healthy adults before they became patients, and we likely have adults in our sample that will go on to become patients. Regardless, our predictions were borne out in the data: the four scores become less correlated at more liberal thresholds, and our chosen threshold seems to be optimal for detecting PRS-AD associations in healthy adults, similar to in patients[20,21]. We added the Discussion:

"We chose a conservative PRS-AD threshold based on studies indicating this shows highest discrimination of patients[20,21], and an assumption that scores would be less comparable at more liberal thresholds, due to including different sets of genetic variants and less consistent effect size estimates (see [22] for why simply deferring to the latest AD GWAS estimates is also not without assumption). Indeed, we found no evidence that incorporating more SNPs in the PRS increased sensitivity to detect genetic effects upon brain ageing. Rather, it may be detrimental to this goal (Supplementary Fig. 4). PRS-AD scores also correlated more poorly when including more variants. The implication is that the choice of GWAS and PRS will affect the outcome of any PRS-AD study, possibly because different AD GWAS capture signals that may become less comparable across the wider genome."



SI Figure 4 (ctd.)
d-e These associations visualized (9 age-subsets on dashed x-axis). The first column depicts the p-values using genome-wide significant SNPs as in the main paper. Coloured points/black outline depicts associations at p<.05 (uncorrected). As these tests are for illustrative purposes only and not independent of our initial main analysis, the FDR-correction level applied is the same as across the 576 PRS-AD tests in the main paper. Dashed and dotted horizontal lines depict uncorrected and corrected significance levels.

2.9. Could the authors add in the methods how to interpret the sign of the 'age relative change' value. That would substantially help the reading.

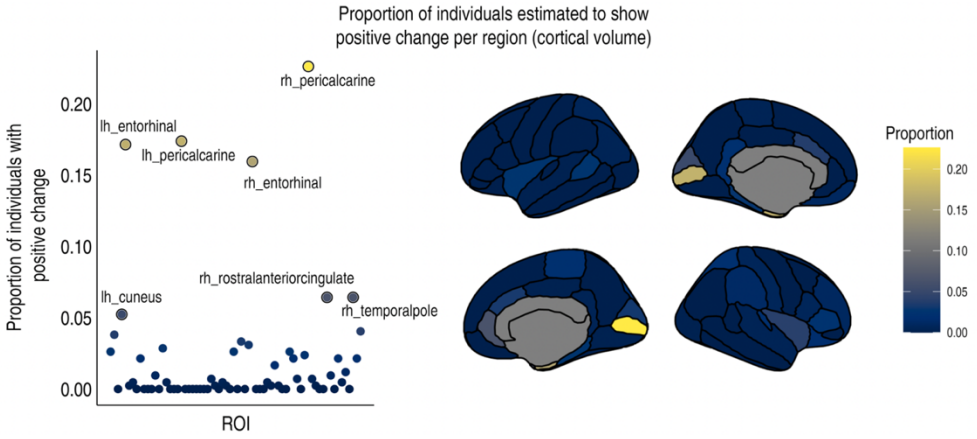
We added the following early in the Results:

"Note that negative absolute change values reflect hippocampal loss, whereas positive would denote an estimated growth. Positive values on age-relative change then correspond to less hippocampal loss than expected given age, whereas negative values reflect more hippocampal loss than expected given age. Fig. 1B-C thus provides the context that higher hippocampal change values correspond to less decline, not growth."

2.10. Also, can they add how to interpret 'absolute change'? Does a positive score in 'absolute change' indicate increased in volume? If so, how do the authors interpret the large number of participants whose entorhinal cortex 'grows' in between the ages of 30 and 70? Is this result plausible, or I am interpreting this metric wrongly?

The Reviewer's interpretation of a positive value on absolute change is correct. The absolute entorhinal change plots in SI Fig 1 imply steep negative slopes in young adults that become less steep by middle age (though still mostly negative), then become steeply negative again around age 60 (SI Fig. 1). The datapoints above 0 in middle age therefore imply estimated growth. We agree this is not plausible, and all other plots imply little evidence for volumetric growth. We thus quantified the proportion of individuals estimated to show volumetric "growth" across the cortex (new SI Fig. 2). This

supports our contention that entorhinal measures are likely among the least reliable. We added the following Discussion:



SI FIGURE 2
Quality control analysis. We estimated the proportion of individuals estimated to show positive absolute change in cortical volume (i.e., "growth") across the cortex. Across cortical regions, the median proportion of individuals estimated to show growth was 0.002. Regions showing >5% individuals with estimated growth are labelled. Left and right entorhinal cortex were clear outliers, suggesting these measures may be less reliable.

"Third, we relied on FreeSurfer-derived measures. While these are well-validated and reliable[7–9], some measures may be less so[9]. Indeed, that we observed no PRS-AD associations with left entorhinal change was surprising. When we quantified the proportion of individuals estimated to show positive absolute change cortex-wide (i.e., "growth"), entorhinal measures were clear outliers, with ~16% estimated to be growing (Supplementary Fig. 2). The median across cortical regions was 0.2%. This may reflect poorer reliability of entorhinal measures, as suggested by others[8,10,11]. Possibly, manual entorhinal tracing or alternative tools may have led to different results[9]."

2.11. Sorry, I could not make sense of Figure 1d with the text provided in the Figure legend or in the main document. The fact that the relative-age change can take positive or negative values with a different meaning (as I understood this metric) makes the interpretation of its association with PRS AD not that straight forwards. I thought more needs to be said about the interpretation of this result.

See previous comment (2.9) for our clarification of this. Further, we made changes to the following sentence and Figure 1 caption:

"The degree of age-relative change was significantly associated with PRS-AD in the hypothesized negative direction: on average across the adult lifespan (30-89 years), individuals losing more hippocampus than expected for their age had significantly higher PRS-AD."

"(Fig. 1) Estimated absolute change per individual (datapoints) in left hippocampus as a function of their mean age (across timepoints). This contextualizes change values in terms of an estimated loss. **c** Estimated age-relative change per individual in left hippocampus (individual-specific slopes) as a function of their mean age. Units are then interpretable in terms of additional hippocampal volume loss per individual, above or below the mean level of loss expected for their age."

2.12. I am puzzled by the function representing absolute change across age for cognition. Can the authors please explain this further and how do they interpret the raise in between 50 and 70 years of age?

This is important to clarify. We added the following Limitations of the mixed-model change estimation, specifically around overinterpreting the estimates, and potential explanations for the shape of this function:

"There are also limitations. First, characteristics of our lifespan data will affect the mixed-model change estimation. Our study had more timepoints in the older age-ranges (Supplementary Fig. 6), likely resulting in more accurate estimates in older adults. Hence, alongside mean age, we corrected for timepoints and the interval between first and last visit to ensure the results were not driven by residual age-related variation. Similarly, normalizing change estimates by age does not reduce variability along the age variable. Because more variability in older age is a known phenomenon[17], adjusting for mean age further ensured the results were not driven by higher dispersion in older adults. Conversely, estimates in younger, less-sampled age-ranges may be

biased by the magnitude of change in older adults. This may help explain why the memory slopes of also younger adults were estimated as negative. Further, selection bias and attrition vary by age-group, which alongside data density differences may explain why adults in their 60's were estimated with less negative memory slopes compared to middle age (Fig. 6B). Caution is thus advised around overinterpreting change estimates in terms of their absolute values, hence we refer to "estimated change"."

2.13. Overall, I thought that the methods applied were too succinctly explained giving their complexity, which will make the understanding of the results obtained difficult to interpret for most readers. I think the manuscript will benefit from more clarity in this respect, considering the wide audience that Nature Communications aims to reach. In general, this manuscript was hard to read, and would benefit from proof-reading.

We have carefully gone through the entire manuscript to improve readability, and have made many changes (see highlighted ms), including:

"To estimate brain change relative to a person's age in adult lifespan data … "

"Change was estimated via the individual-specific random slopes in a Generalized Additive Mixed Model (GAMM) of age (sex, scanner, and intracranial volume [ICV] corrected)."

"Then, in 364 features we modelled a GAMM of age (irrespective of group), and entered the individual-specific slopes into ML binary classification (Fig. 3B; sex, scanner field strength, and ICV corrected; note that the ADNI sample did not change field strength over time)"

"In the LCBC healthy adult lifespan discovery sample, we then modelled adult lifespan change in all 364 features used to train the AD-control model, and estimated the individual-specific slopes as before. Then, we calculated a multivariate marker of change based on the list of features the model found most important for classifying AD patients from controls, and related this to PRS-AD."

"Finally, in the LCBC adult lifespan discovery sample, we separated individuals into discrete groups based on the conjunction of brain and genetic risk factors. For this, we used the partial association between PC1 of age-relative change calculated across the first 50 AD-sensitive features – here including hippocampal and amygdala volumes ($PC1^{relChange1\text{-}50}$) – and PC1 calculated across the four PRS-AD scores ($PC1^{PRS\text{-}AD}$; explaining 87%; Methods). We hypothesized that high PRS-AD individuals who are also high on a multivariate marker of change in AD-sensitive features would show more longitudinal memory decline (pink quadrant 4 in Fig. 6D; Methods). Akin to the brain analysis, memory-change estimates were derived via the individual-specific random slopes in a GAMM of age"

2.14. "Genetic AD risk is robustly associated with the slope of brain ageing in healthy adults" -> I feel this is an overstatement. It is true that the authors provide some evidence for this, but I would not call that 'robust'. The authors certainly report results convincingly associating AD PRS with hippocampal volume; however, when APOE-e4 carrier status is accounted for, the associations clearly drop, and no interaction agexAD-PRS is found to survive multiple testing in the full-age range analysis. Results, become far less consistent for their analyses of ROIs for Braak stages I and III.

We changed the opening statement of the Discussion to:

"Variation in brain ageing trajectories in healthy adults links with AD-related genetic variation and memory decline outcomes through adulthood."

We also made it even more clear that the associations dropped after excluding APOE, which we note we transparently visualize in our figures:

"These associations with multivariate change measures were largely though not entirely driven by *APOE"*

"Of note, while PRS-AD effects were not solely driven by *APOE*, *APOE* nevertheless accounted for much of the predictive power of PRS-AD, as associations often disappeared or were attenuated using $PRS\text{-}AD^{noAPOE}$. This fits with studies finding PRS-AD associations with cognitive and metabolic factors in adults are largely driven by *APOE*[23], and limited utility of SNP's beyond *APOE* to predict AD markers[24]. Most associations after excluding *APOE* were with scores derived from the genome-wide significant SNP's/weightings reported by Jansen et al.[4], possibly suggesting these better capture differences in brain ageing (though $PRS\text{-}AD^{noAPOE}$ effects were also evident using scores from two other GWAS[5,6])."

Our main strength was in the longitudinal sampling, enabling us to estimate subject-specific change. The agexAD-PRS interaction results are presented as an alternative analysis approach and provided in the supplement for the sake of completion. However, this never constituted our approach, because we knew that despite our dense longitudinal sample, we would lack power across the full age-range to

detect a 2-way continuous interaction with derived change values. We therefore used an alternative approach inspired by Specification Curve Analysis, and to correct for comparisons across tests. We agree the univariate results for Braak stages I and III are somewhat inconsistent. Our paper may thus suggest there are limitations to univariate measures. When we adopted a multivariate approach, the results across structures became more consistent (Fig. 4). We thus added the following, and also discuss the benefits of our approach over single-shot measures of statistical significance:

"Relatedly, our results point to the advantage of multivariate measures of change over univariate measures."

"As inferences based on significance are affected by arbitrary analysis choices, we took inspiration from multiverse methods to define a defensible set of choices to perform analyses across[25,26]. In our case, the main arbitrary covariate was the age-range to test the association across. Despite accounting for age- and time-related covariates, the influence of this choice on statistical significance is clear in Fig. 2, Fig. 4B and Fig. 5. This clarifies why we used multiple scores; using a single PRS could have obscured the results, as significance fluctuated across scores, or using the same score across age-range specifications. Adopting this approach, we could ensure capture of ageing-specific processes, document the stability of PRS-AD-change associations in healthy adults, and ensure the results were independent of a single arbitrary decision[25,26]"

2.15. "By specifically isolating within-individual genetic effects on accelerated brain ageing, the present study confirms AD risk genes also influence normal variation in hippocampal change rates in healthy adults." I don't think this conclusion is correct, as far as I understood the authors did not limit their polygenic scores to SNPs within risk genes, or even exomes; so, I don't think any conclusion can be drawn with regards of where this signal is coming from within the genome (may well be driven by SNPs outside genic regions).

Sorry this was not clear. Our intention was not to say where in the genome the signal arises but that our longitudinal approach specifically addressed the genetic effect upon change, now amended to:

"And since AD risk variants may influence hippocampal differences early in life[27–29], cross-sectional findings in older adults[30–32] cannot attribute genetic effects to accelerated brain ageing[33]. By isolating genetic effects on change, our study confirms AD risk variants influence variation in hippocampal change rates in healthy adults."

2.16. I found very little justification as per why the authors applied other p-thresholds than the genome-wide originally selected throughout the reading of the results and discussion. As mentioned earlier, I do not really understand why to use such stringent p-threshold when calculating a PRS, since the authors are rejecting a lot of valid information; and I did not find in the paper any convincing argument to such unusual strategy. Likewise, there is no clear justification or interpretation for the use of 4 different GWAS results.

Please see our previous responses (2.6 & 2.8) and SI Fig 4. Our predictions were supported by the data. See 2.14 for text justifying our use of multiple scores.

2.17. On a side note, were the participants included in this study, also included in any of those GWAS?

None of the adult lifespan samples used to test PRS-AD associations were included in these GWAS.

2.18. In general, I found the discussion overstated and would suggest downplaying the conclusions of this study.

We made many amendments to the Discussion (see ms) and the Conclusion (below). Thank you for your constructive and insightful comments.

"In conclusion, brain ageing trajectories in healthy adults are accelerated by the presence of AD risk variants, in many brain features, and in some cases beyond *APOE*. We show that brain features most susceptible to faster deterioration in AD are on a trajectory of accelerated change from age ~50 in healthy individuals, and that models trained on AD patients can be applied to adult lifespan data and the prediction relates to genetic AD risk. Finally, genetically at-risk individuals with more brain change showed more adult lifespan memory decline, compared to genetically at-risk individuals with less brain change. Thus, brain change in AD-sensitive regions enhanced the value of knowing a person's genetic risk, and atrophy predicted memory decline more than the genetics. Our findings show that brain ageing slopes in healthy adults correlate with AD-related genetic variation and memory decline through adulthood, and that neurodegeneration occurs along a continuum from normal ageing to AD – supporting a dimensional, lifespan view of AD."

Reviewer #3 (Remarks to the Author):

This is a very impressive article that asks some fundamental questions about trajectories of structural decline in healthy adults compared to participants with AD from the Alzheimer's Disease Neuroimaging Initiative (ADNI). The authors used machine learning to characterize brain regions that declined most steeply with AD. Then they isolated an adult lifespan sample of normal adults aged 30-90 and assigned each subject a "polygenic risk score (PRS)" based on APOe4 status and other genetic markers of AD. This set the stage to allow them to test, using multiple methods, whether those with higher risk scores showed a pattern of volume loss similar to the AD patients. Using a sliding window that progressively narrowed their sample by age, they determined that an accelerated pattern of change similar to AD patients typified those with higher PRS scores. The analyses are impeccable, thorough, and I would actually describe them as elegant.

We thank the Reviewer for their positive feedback and helpful comments.

3.1. I think this study unquestionably validates the use of PRS scores as measures of risk. But that is not really the aim of the study. The introduction focuses on whether normal aging and AD follow similar trajectories but at different rates. It seems to me that the theoretical and empirical work that would be most informative would be to use the powerful and creative methods described to determine the youngest age at which a predictive pattern of polygenic factors could validates the use of polygenic markers in assessing risk, but this be found. As it stands now, by selectively decreasing the sample from youngest to oldest, the authors show that the pattern that is characteristic of older adults is similar to AD patients. The sliding window approach and accompanying methods seem ideal for examining this issue and could readily be adapted. The findings would be hugely interesting but require a true lifespan sample and a much larger n.

We agree this would be a great application of our method, but that the suggested analyses would likely require a much larger N for the genetic sample. We included SI Figure 6 showing the age distributions of the MRI dataset and its genetic subset, and added text to outline the uneven age distribution (see response 2.2). Still, we carried out the suggested analysis examining the parameter estimates for PRS-AD associations after discarding the data from the oldest individuals. While we cannot draw strong conclusions due to the differences in genetic sample size between age-ranges (see below Fig), the data suggest the associations are not solely driven by the oldest participants. Note that all analyses correct for age, N timepoints, and the interval between first and last visit. We agree we mainly show the pattern characteristic of older adults is similar to AD, and have tempered conclusions pertaining to lifespan ageing where it is not supported by the data.

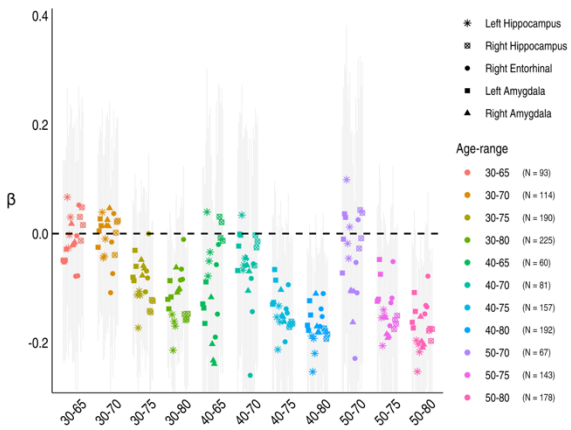
Yet we also ran the suggested analysis for memory decline. As we expected, this revealed we can detect significant associations with memory decline using brain and memory change estimates in fairly young individuals (e.g., within the age-range 30-65). Again, all analyses correct for age++. We added Supplementary figure 17, which extensively reports the memory decline results tested in different age subsets (including all N's). We also added the following in the Results and Discussion. Note we refer to the memory decline results as reflecting memory decline "through adulthood", which is supported by the data:

"Finally, the group difference in memory decline was not driven by the oldest adults or by residual group-differences in age, but persisted using change estimates from alternative age subsets (see Supplementary Fig. 17)"

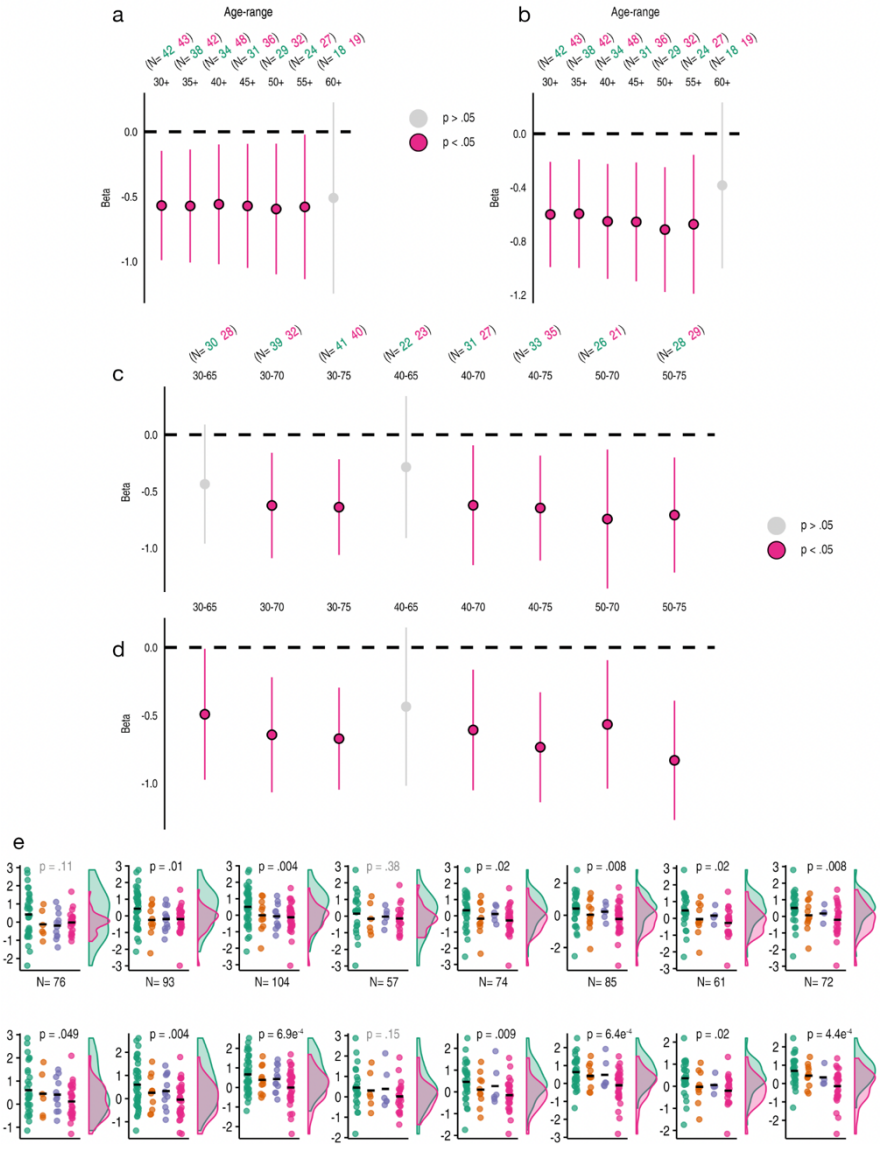
"Moreover, group differences in memory decline were more driven by brain change differences than by genetic differences, as they persisted when controlling for PRS-AD and *APOE-ε4* but not atrophy. Thus, brain change may be crucial for detecting comparatively at-risk individuals in adult lifespan data. Further, memory decline differences were protracted through adulthood, as they were evident in different age subsets, including comparatively younger adults (e.g., within the age-range 30-65; Supplementary Fig. 17). This indicates neurodegeneration in AD-sensitive regions tracks with memory decline differences that are detectable through adulthood. It also emphasizes that memory decline is a gradual phenomenon not confined to old age. This perspective may be obscured in studies that use clinical tests that do not capture subtle cognitive variations, and estimates of decline relative to a group rather than one's earlier capacity."

"Finally, genetically at-risk individuals with more brain change showed more memory decline through adulthood, compared to genetically at-risk individuals with less brain change. Thus, tracking change in AD-sensitive regions enhanced the value of knowing a person's genetic risk, and atrophy predicted memory decline more than the genetics. Our findings show that brain

ageing slopes in healthy adults correlate with AD-related genetic variation and memory decline through adulthood, and that neurodegeneration occurs along a continuum from normal ageing to AD."



SI FIGURE 6
Beta estimates for PRS-AD associations with age-relative change in early Braak stage regions, tested across alternative age subsets discarding the data from the oldest individuals.



SI FIGURE 17
Memory change results for models across different age-ranges. a Group-differences in age-relative and b absolute memory change tested at progressively older age-ranges (main model). c Group-differences in age-relative and d absolute memory change across alternative age subsets (redundant subsets removed; standardized betas, 95% CI error bars). Colour denotes significance at $p < .05$ [uncorrected]. As these tests shed light on our main hypothesis-driven model, no correction was applied. e The distribution of group differences in age-relative (upper row) and absolute memory change (lower) within each age subset (Y axis shows Z values). The p-value of the group-difference is given above each plot. The reported group differences in memory change were not driven by the oldest adults but evident within the age-ranges 30-65 and 30-70 (all models and datapoints corrected for covariates: mean age, APOE-ε4 carriership, sex, N timepoints, and interval between first and last timepoint). The N of the two groups wherein the difference was found (a-d) and N of the overall model is given (e).

3.2. Critically, it is a truly unfortunate weakness that the present study, overall has a small number of subjects for a lifespan study and the number of 40-50 year olds is sparse—it appears to comprise only a fraction of the lifespan sample. In general, middle aged individuals are severely under sampled.

We agree it is unfortunate that the genetic sample is small. However, the lifespan sample used to estimate brain change is one of the most densely sampled longitudinal lifespan brain studies to date (from a single site). We thus find it encouraging that the methods we use to estimate subject-specific change lead to consistent results even in small genetic samples. Middle adulthood is comparatively less sampled due to higher naturally occurring dropout in longitudinal studies in this age-range (see also 3.4). We added the following in Limitations:

"Second, we used all available longitudinal data to estimate brain and memory change, each in a single model. This likely optimized the change estimates for all, including the relatively modest subset with genetic data"

"… characteristics of lifespan data will affect the mixed-model change estimation. Our study had more timepoints in the older age-ranges (Supplementary Fig. 6), likely resulting in more accurate estimates in older adults. Hence, alongside mean age, we corrected for timepoints and the interval between first and last visit to ensure the results were not driven by residual age-related variation. Similarly, normalizing change estimates by age does not reduce variability along the age variable. Because more variability in older age is a known phenomenon[17], adjusting for mean age in all association tests further ensured the results were not driven by higher dispersion in older adults. Conversely, estimates in younger, less-sampled age-ranges may be biased by the magnitude of change in older adults. This may help explain why the memory slopes of also younger adults were estimated as negative. Further, selection bias and attrition vary by age-group, which alongside data density differences may explain why adults in their 60's were estimated with less negative memory slopes compared to middle age (Fig. 6B). Caution is thus advised around interpreting change estimates in terms of their absolute values, hence we refer to "estimated change"

3.3. I have been considering what pattern of findings might have occurred that would be unexpected. To the extent that the PRS is valid, the result seem preordained. No one comes down with AD acutely and given that a large number of older adults would likely have incipient AD, it doesn't seem that surprising that they have volume loss and decline rates that mimic AD. It is important to keep in mind that adults who have AD or will get it are included in the normal sample. I believe that a study of low amyloid, low PRS adults is necessary to understand what typifies the aging process and what is shared by normal aging and AD. Such a study would likely lead to a large and interesting debate about what constitutes normal aging.

We agree that despite excluding based on cognitive and depression scores to guard against including participants with incipient AD, we cannot exclude this possibility, and some older adults here may have biomarkers or be diagnosed with AD later in life. While we note that mean atrophy patterns of low genetic risk-low amyloid participants have also been shown to be largely overlapping with AD (Fjell et al., 2013), future research with biomarkers will be required to rule out whether the neurodegenerative patterns we see that relate to memory decline are driven by unmeasured AD biomarkers. We added the following Limitation:

Fjell, A.M., McEvoy, L., Holland, D., Dale, A.M., Walhovd, K.B. and Alzheimer's Disease Neuroimaging Initiative, 2013. Brain changes in older adults at very low risk for Alzheimer's disease. *Journal of Neuroscience*, *33*(19)

"Sixth, we used only structural measures. While these are sensitive to detecting subtle changes in brain structure that ultimately reflect a continuous, lifelong process of change, other biomarkers are necessary to refine detection of AD-risk in healthy adult samples. Finally, we do not know which individuals here will be diagnosed with AD later in life, or have other AD biomarkers suggesting a biological trajectory to AD[2]. While our analyses suggest one could assign differential transition probabilities to healthy individuals, only time and follow-up data will tell."

Note also that our dataset comprises adults who are unrepresentatively high performing, likely due to demands involved with longitudinal studies that lead to participation bias and selective attrition. We added text emphasizing the cognitive scores of our sample (along with SI Fig. 18 and SI Table 2). We also highlight that our study likely captures variation in brain and cognitive trajectories in adults within the cognitively healthy range that imply importance for AD-related outcomes.

"Generally, the LCBC sample is comprised of cognitively high-performing individuals (see summary of cognitive scores in Supplementary Table 2; Supplementary Figure 18)"

"However, that study recruited participants with memory complaints and a family AD history via memory clinics. In contrast, our sample comprised healthy adults in longitudinal studies which are known to be biased towards retaining high performers [34,35] (as may be evident in the cognitive scores of the older adults here; see Supplementary Table 2; Supplementary Fig. 18)."

"Seventh, longitudinal studies inevitably culminate in unrepresentatively high-performing samples[58]. Since even in these we find variation in brain ageing slopes that correlates with AD-related genetic variation and memory decline, the population effect-sizes may be larger."

"This indicates neurodegeneration in AD-sensitive regions tracks with memory decline differences that are detectable through adulthood. It also emphasizes that memory decline is a gradual phenomenon not confined to old age. This perspective may be obscured in studies that use clinical tests that do not capture subtle cognitive variations, and estimates of decline relative to a group rather than one's earlier capacity."

"Our findings show that brain ageing slopes in healthy adults correlate with AD-related genetic variation and memory decline through adulthood, and that neurodegeneration occurs along a continuum from normal ageing to AD – supporting a dimensional, lifespan view of AD."

3.4. A notable strength of the paper is the combining of different brain structures and examining whether the findings continue to be validated when multiple combinations of structures are tested. Similarly, the finding of more accelerated memory decline coupling with vulnerable, high PRS subjects is another strength. I find this to be an outstanding, landmark study that besides being very informative about AD and some aspects of its etiology, also provides an important template for methods of combining multiple methods and a diversity of analytic techniques to answer some broad questions. of machine language techniques further strengthens the paper. My only reservation is that I think that the conclusions should not be overgeneralized to lifespan aqing, given the limited sample of middle-aged adults and I think this should be explicitly noted as a weakness. The authors also correctly note the limitations of not having measures of amyloid available.

We have noted the uneven timepoint distribution as a weakness (see 3.2) and tempered our conclusions pertaining to lifespan ageing accordingly. See also SI Fig 6 (response 2.2) for the timepoint distribution across age. While we do have less observations from middle age compared to older age, 456 longitudinal brain observations come from 170 individuals with a mean age < 60, 253 from 92 with a mean age < 50, and 159 from 56 with a mean age < 40. Thus, while our longitudinal coverage in older adults is better, the coverage in middle age is arguably not poor.

3.5. Despite a few concerns. I find this to be an outstanding, landmark study that besides being very informative about AD and some aspects of its etiology, also provides an important template for methods of combining multiple data collection techniques with a diversity of analytic techniques to answer some broad questions. There is unusually good evidence´for replicability ranging from multiple regression to machine learning. The study blends a cognitive aging approach to a more traditional medical sciences approach. I believe we will only solve the AD problem if these groups join forces, and this collaboration clearly shows how to do it. I do recommend that the uneven age sampling, particularly middle-aged, should be explicitly noted as a limitation in making general comments about aging. The authors also correctly note the limitations of not having measures of amyloid. This study gave me great hope and optimism about the study of AD and aging. Regardless of the editorial decision, the authors are to be congratulated
For their forward thinking and excellent model for achieving scientific advances.

Thank you for your valuable insight, kind words, and for helping refine our study. We added the limitations as suggested (3.4).

Reviewer #4 (Remarks to the Author):

This research aims to investigate the effect of genetic Alzheimer's disease (AD) risk on brain changes during healthy aging. Motivated by the qualitative similarities between healthy brain aging and AD, the authors hypothesize that a higher genetic risk for AD may contribute to faster atrophy rates in healthy adults. Their literature review indicates that previous studies have predominantly relied on cross-sectional designs, resulting in mixed results. Additionally, the reviewed longitudinal studies failed to account for individual differences in brain features. To address these limitations, the authors propose an approach to compute individualized

estimates of brain atrophy relative to age. First, they examine genetic AD risk in relation to early Braak regions, known to be affected in both healthy aging and AD. Then, using a data-driven method, they identify additional brain regions exhibiting accelerated atrophy in AD and analyze their associations with genetic AD risk. Finally, this research explores genetic AD risk in relation to individualized estimates of memory change over the course of aging in healthy adults.

Key results:
In a cohort of healthy adults, polygenic risk scores for AD (PRS-AD) are associated with reductions in key brain features affected in AD, specifically hippocampal volume, and a data-driven multivariate marker of brain change. Interestingly, these associations are largely, though not exclusively, driven by APOE. Finally, among healthy adults who are at high genetic risk for AD, those with higher rates of brain change exhibit greater memory decline relative to those with less brain change.

Validity
The claimed results of the study are:
1. PRS-AD are associated with greater age-relative change in early Braak regions (the hippocampus, amygdala, and entorhinal cortex). These extend beyond the contribution of APOE.
2. a) PRS-AD is significantly associated with change in a multivariate marker of AD-accelerated brain features in healthy adults.
b) Some of the PRS-AD associations with brain change in healthy adults are replicated in an independent sample.
3. Among individuals with high PRS-AD, those who are high on the multivariate marker of brain change present greater reductions in memory over the adult lifespan relative to those with lower brain change.

Claim 1: This result is most convincing in the left hippocampus, where multiple PRS-AD scores are associated with age-relative change and absolute brain change for most of the age-ranges. 86% of these survive FDR-correction and 33% remain significant when APOE is taken out of the PRS-AD. The associations with the right hippocampus and right amygdala are largely driven by APOE. The associations with the entorhinal cortex and the cortical ROIs of Braak Stage III are much weaker/non-existent after FDR correction.

While we agree the associations with right hippocampus were largely driven by APOE, we note we nevertheless found some evidence for PRS-AD^{noAPOE} associations, and PRS-AD^{noAPOE} associations were confirmed in the replication sample.

Claim 2a: 14 of the 36 tested associations relating the multivariate marker of brain change to PRS-AD were significant after FDR-correction, but the results are largely driven by APOE (one association remains significant without APOE, in the age group 65+) It is interesting that the sliding-window PCA yielded similar results, ensuring that no specific brain regions are driving the multivariate effect. It was thoughtful to run the analyses with and without APOE, as well as with a sliding window PCA as this facilitates interpretation.

Claim 2b: The association with the left hippocampus is replicated, and age ~50 is again found to be the time when brain changes increase. However, the PRS-AD associations with the amygdala, and the multivariate marker of change are not replicated. The latter is a key finding in the study, and the lack of replication is an important limitation. The authors mention the replication sample not having as many follow-ups, but elaborating on other explanations would be helpful.

We added the following in the limitations, also in response to your later comment concerning APOE:

"Fourth, while the discovery sample screened out participants with mental disorders, the replication sample included individuals with disorders (Methods). This decision aimed to increase power to estimate change in the less-powered replication sample, but potentially influenced the results."

"Sixth, alongside the more limited longitudinal coverage which will negatively impact change estimates, sampled or geographic differences in *APOE* genotype may account for the lack of full replication in the independent adult lifespan cohort (Supplementary Figure 19; Supplementary Table 8)"

Claim 3: This finding is interesting, especially because it is not driven by APOE e4 carriership. It suggests that individuals with both a higher genetic risk for AD, and a higher index of brain change are more likely to exhibit greater memory decline over the course of aging. The analyses, however, also suggest the brain change marker has more predictive power than PRS-AD, thus undermining the relevance of looking at genetic AD risk.

We agree, and have made this point more explicit in the Discussion and Conclusion:

"Moreover, group differences in memory decline were more driven by brain change differences than by genetic differences, as they persisted when controlling for PRS-AD and *APOE*ε4 but not atrophy. This suggests brain change may be crucial for detecting comparatively at-risk individuals in adult lifespan data."

"Thus, tracking change in AD-sensitive regions enhanced the value of knowing a person's genetic risk, and atrophy predicted memory decline more than the genetics."

Significance
This research highlights the parallels between healthy aging and AD and suggests that neurodegeneration occurs along a continuum. This holds clinical significance, especially given the growing interest in combination therapies that target the biology of aging in AD. The finding in relation to memory decline further emphasizes the translational significance of the paper. Finally, the findings add to the lifespan neuroscience literature while addressing methodological gaps.

Data and methodology
The data samples appear to have been carefully selected and preprocessed.

Analytical approach
The Generalized Additive Mixed Model is used to compute individualized estimates of change in brain features over aging. The specified models are reasonable, and this approach allows for individualized estimates of brain change over time. The XGBoost binary classifier is trained on brain change features to differentiate those with an AD diagnosis at the final timepoint from those with normal cognition across all time points. The model is validated externally on AIBL. The approach is reasonable, and the chosen model seems to have traded specificity for sensitivity.

We have added text to explain why the classifier had high sensitivity but poor specificity:

"Importantly, because the model was trained on an index of relative brain change conditioned on age, the logistic prediction applied to the healthy adult lifespan data cannot be interpreted in terms of its implied binary outcome (i.e., AD/no-AD). This is because the model could assign the same probability of having AD to a hypothetical 30-year-old with an estimated additional brain loss of 10mm^3/year as to a 60-year-old with the same additional brain loss. Despite the change being higher in the 60-year-old, because it exceeds the mean brain loss anticipated at age 60 (see Fig. 1C). Note that this clarifies why the model had high sensitivity but lower specificity in separating AD from controls in AIBL; it is a characteristic of the measures we used for model training more than the model itself."

Suggested improvements
4.1. Figure 2, 4, 5: it is difficult to interpret what it means for one PRS-AD score to remain significantly associated with brain change, while another does not after FDR correction and/or removal of APOE. A brief explanation of the four summary scores, and how they differ from one another would be helpful in the introduction.

While we agree this is not very intuitive, it does highlight the advantage of our multiverse-inspired approach over single-shot tests of significance. We added the following in the Discussion:

"As inferences based on significance are affected by arbitrary analysis choices, we took inspiration from multiverse methods to define a defensible set of choices to perform analyses across[25,26]. In our case, the main arbitrary covariate was the age-range to test the association across. Despite accounting for age- and time-related covariates, the influence of this choice on statistical significance is clear in Fig. 2, Fig. 4B and Fig. 5. This clarifies why we used multiple scores; using a single PRS could have obscured the results, as significance fluctuated across scores, or using the same score across age-range specifications.

Adopting this approach, we could ensure capture of ageing-specific processes, document the stability of PRS-AD-change associations in healthy adults, and ensure the results were independent of a single arbitrary decision[25,26]"

We also added analyses showing the correlation between PRS-AD scores at different p-value thresholds, which shows highest correlations between scores at our chosen threshold:

"… at more liberal SNP inclusion thresholds, the four scores became less comparable, dropping from a median correlation between scores of R = .73 at our chosen threshold, to R = .29 at the most liberal threshold"

Thus, our results suggest the choice of PRS score and GWAS are critical decisions likely to impact the results of a PRS-AD study. We added Discussion around this:

"We chose a conservative PRS-AD threshold based on studies indicating this shows highest discrimination of patients[20,21], and an assumption that scores would be less comparable at more liberal thresholds, due to including different sets of genetic variants and less consistent effect size estimates (see [22] for why simply deferring to the latest AD GWAS estimates is also not without assumption). Indeed, we found no evidence that incorporating more SNPs in the PRS increased sensitivity to detect genetic effects upon brain ageing. Rather, it may be detrimental to this goal (Supplementary Fig. 4). PRS-AD scores also correlated more poorly when including more variants. The implication is that the choice of GWAS and PRS will affect the outcome of any PRS-AD study, possibly because different AD GWAS capture signals that may become less comparable across the wider genome."

4.2. Authors could explore additional modeling architectures in addition to XGBoost. This would help ensure that the findings are consistent and relatively agnostic to model frameworks.

Respectfully, the machine learning (ML) methods are not the focus of this paper. We show our model performs well in independent AD-control data, and that when we apply the ADNI-derived weights to adult lifespan change estimates relative to age, the prediction relates to PRS-AD in healthy adults. Thus, the model has been validated in not one but two independent datasets. We agree that an avenue for future research will be to refine the ML prediction methods, particularly when applying AD-derived models to healthy adults. Our paper serves as proof-of-principle this is possible, but through an ML approach that was guided by a theoretical understanding of brain changes in ageing and AD.

4.3. Additional performance metrics should be reported, such as area under the precision-recall curve. The mean results over the 500 iterations of the 10-fold cross-validation should be reported as well.

We added this in Results and Methods:

"precision-recall AUC = .883"

"we selected the final hyperparameters based on the mean AUC obtained across the 500 iterations of 10-fold cross-validation (mean = 0.927), where each iteration logged the maximum AUC achieved across folds"

4.4. The lack of association with the entorhinal cortex is surprising. All brain features were measured via FreeSurfer, which, according to the authors, limit accurate measurements of the entorhinal cortex. They mention manual tracing as a potential solution, but this may not be feasible in a large multi-cohort dataset. Alternative solutions could include other processing toolboxes (such as CAT12) and atlases (such as Hammers_mith) or extracting radiomic features, and then comparing the results with the FreeSurfer measures.

We ran QC analyses that suggest the FS-derived entorhinal measures are likely among the least reliable cortical measures, adding the following text and SI Fig 2 (please see also response 2.10). However, I'm sure the Reviewer agrees we are no longer at liberty to explore alternative measures because the results for this structure were not as expected. Rather, it is important to present the results as they are and allow researchers to judge their likelihood for themselves. In future projects, we will certainly explore alternative entorhinal measures.

"Third, we relied on FreeSurfer-derived measures. While these are well-validated and reliable[7–9], some measures may be less so[9]. Indeed, that we observed no PRS-AD associations with left entorhinal change was surprising. When we quantified the proportion of individuals estimated to show positive absolute change cortex-wide (i.e., "growth"), entorhinal measures were clear outliers, with ~16% estimated to be growing (Supplementary Fig. 2). The median across cortical regions was 0.2%. This

may reflect poorer reliability of entorhinal measures, as suggested by others[8,10,11]. Possibly, manual entorhinal tracing or alternative tools may have led to different results[9]"

4.5. The authors should discuss the potential effects of scanner type, field strength and sequence parameters.

We have clarified that our "ADNI sample did not change field strength over time" (see also response 1.1). The main LCBC sample did, and we agree this will influence variability in the estimates. We thus added the Limitation:

"Relatedly, scanner parameters changed over time for some samples (Supplementary Table 4). While we made efforts to correct for or reduce scanner variation (Methods), this will influence estimates."

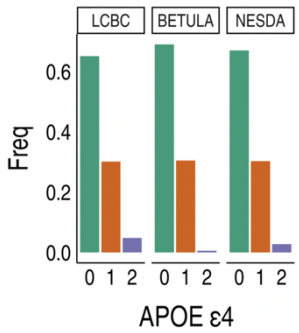
4.6. The sample's breakdown by race and ethnicity is not reported; if the sample is homogenous, it should be mentioned as a limitation in the discussion.

We do not have these data, so we added the following Limitation:

"Fifth, the adult samples consisted mainly of homogenous white ethnic populations from their respective countries, as did the GWAS on which PRS scores were based, possibly limiting result generalizability."

4.7. Regarding the replication analyses: since APOE seems to be the primary factor driving the associations between genetic AD risk and brain change, it is possible these associations are not significant in the replication sample if it happens to have fewer participants who are APOE-e4 carriers. It could be helpful to look at the distribution of the APOE genotype in the replication sample to help explain the lack of replication.

We added description of APOE genotypes in SI Table 8, SI Figure 19, and the following Limitation:



SI FIGURE 19
Frequency of *APOE* ε4 genotype per sample. Fisher's exact tests indicated there were no significant differences between samples in the number of ε4 carriers (all p > .45). A significant difference in APOE ε4 genotype was found between LCBC and BETULA only (p = .03), likely reflecting more ε4 homozygotes in LCBC (see SI Table 8).

"Sixth … sampled or geographic differences in *APOE* genotype may account for the lack of full replication in the independent adult lifespan cohort (Supplementary Figure 19; Supplementary Table 8)"

4.8. The clinical implications of the study could be clarified.

We added text in Discussion to reflect possible clinical implications, and changed sentences in the Abstract and Conclusion to highlight that our study captures continuous variation in AD-related risk factors in healthy adults:

"Whether these brain trajectory differences are relevant for later AD outcomes will require follow-up and biomarker assessment, but our results indicate these neurodegenerative changes are not benign. They also underscore the need for follow-up data over extended age-spans for prediction or prevention of AD, and suggest a continuous view on lifespan brain health may aid understanding of AD[78,79]. Multivariate atrophy measures may help assess AD risk and improve selection into clinical trials."

(Abstract) “Our findings suggest that quantitative AD risk factors are detectable in healthy individuals, via a shared pattern of ageing- and AD-related neurodegeneration that occurs along a continuum and tracks memory decline through adulthood.”

“Our findings show that brain ageing slopes in healthy adults correlate with AD-related genetic variation and memory decline through adulthood, and that neurodegeneration occurs along a continuum from normal ageing to AD.”

4.8a. Do the results suggest that healthy individuals with higher PRS-AD and higher age-relative brain changes may develop mild cognitive impairment or dementia in the future? Is it possible these clinically healthy individuals may already be on the biological AD trajectory? ADNI and AIBL have biomarker data, which could be used to gain insight into the mechanisms underlying the observed associations.

We cannot know from this study, which researched mostly high-performing individuals in longitudinal studies but lacked biomarkers. We agree this is a critical follow-up study. In addition to the above, we thus added:

“Finally, we do not know which individuals here will be diagnosed with AD later in life, or have other AD biomarkers suggesting a biological trajectory to AD[2]. While our analyses suggest one could assign differential transition probabilities to healthy individuals, only time and follow-up data will tell.”

4.8b. Is it only fair to say that individuals with greater age-relative changes in brain volumes are likely to have a higher PRS-AD, or does this association mean these individuals are likely to develop dementia? The authors touch on this limitation at the end of the discussion, but it could be helpful to clarify the implications of the study despite this limitation.

In addition to the above (4.8a), see the highlighted amendments to the abstract and conclusion (in 4.8) for how we frame our findings.

4.8c. The authors mention that they disregarded AD heterogeneity as this was reasonable for their purpose of identifying brain features that tend to change faster in AD. However, it is possible that the “age-relative change” metric is capturing changes driven by different subtypes rather than by individual differences. It could be helpful to run additional analyses accounting for the major AD subtypes.

We agree it is possible our metric may be capturing changes partly reflecting heterogeneous atrophy patterns or subtypes. It remains a discussion whether AD subtypes are discrete or continuous in nature (Mohanty et al., 2023), but it nevertheless seems that the majority of individuals are best described by an AD-typical pattern, with example estimates of between 33% (Vogel et al., 2023) and 55% (Ferreira et al., 2020) falling into the AD-typical class described by the Braak stages. Further, ageing research indicates the variance in longitudinal brain change is captured by one major dimension of atrophy (~66%; Cox et al., 2021). Given the modesty of our genetic sample size, it would not be practical for our study to attempt to partition variance in change in order to account for patterns that are expected to be present in a minority of individuals at best. We amended the Discussion of subtypes to the following. Thank you for your insightful comments.

Mohanty, R. et al., 2023. Associations between different tau-PET patterns and longitudinal atrophy in the Alzheimer’s disease continuum: biological and methodological perspectives from disease heterogeneity. Alzheimer's Research & Therapy, 15(1).

Vogel, J.W. et al., 2021. Four distinct trajectories of tau deposition identified in Alzheimer’s disease. Nature medicine, 27(5).

Ferreira, D., Nordberg, A. and Westman, E., 2020. Biological subtypes of Alzheimer disease: a systematic review and meta-analysis. Neurology, 94(10).

Cox, S.R. et al., 2021. Three major dimensions of human brain cortical ageing in relation to cognitive decline across the eighth decade of life. Molecular Psychiatry, 26(6).

“Second, our approach disregards heterogeneity in ageing or AD-related atrophy, treating all individuals with an AD diagnosis as one group compared to all normal controls. This was reasonable for our goal of identifying features with faster average change in AD, given there may be a predominant AD atrophy pattern[36] that overlaps with the average ageing pattern[37–39]. But since there are AD subtypes[36,40,41], an important question is whether AD variability traces to brain change heterogeneity in adults.”

Reviewer #5 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports as part of the Nature Communications initiative to facilitate training in peer review and appropriate recognition for co-reviewers

Thank you.

1. Chow, N. *et al.* Comparing 3T and 1.5T MRI for mapping hippocampal atrophy in the Alzheimer's disease neuroimaging initiative. *Am. J. Neuroradiol.* **36**, 653–660 (2015).
2. Ossenkoppele, R. *et al.* Amyloid and tau PET-positive cognitively unimpaired individuals are at high risk for future cognitive decline. *Nat. Med.* **28**, 2381–2387 (2022).
3. Brouwer, R. M. *et al.* Genetic variants associated with longitudinal changes in brain structure across the lifespan. *Nat. Neurosci.* **25**, 421–432 (2022).
4. Jansen, I. E. *et al.* Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat. Genet.* **51**, 404–413 (2019).
5. Lambert, J. C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* **45**, 1452–1458 (2013).
6. Kunkle, B. W. *et al.* Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. *Nat. Genet.* **51**, 414–430 (2019).
7. Cardinale, F. *et al.* Validation of FreeSurfer-Estimated Brain Cortical Thickness: Comparison with Histologic Measurements. *Neuroinformatics* **12**, 535–542 (2014).
8. Han, X. *et al.* Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer. *Neuroimage* **32**, 180–94 (2006).
9. Leng, Y., Ng, K. E. T., Vogrin, S. J., Meade, C. & Ngo, M. Comparative Utility of Manual versus Automated Segmentation of Hippocampus and Entorhinal Cortex Volumes in a Memory Clinic Sample. **68**, 159–171 (2019).
10. Hedges, E. P. *et al.* Reliability of structural MRI measurements: The effects of scan session, head tilt, inter-scan interval, acquisition sequence, FreeSurfer version and processing stream. *Neuroimage* **246**, 118751 (2022).
11. McGuire, S. A. *et al.* Reproducibility of quantitative structural and physiological MRI measurements. *Brain Behav.* **7**, 1–17 (2017).
12. Marioni, R. E. *et al.* Genetic Stratification to Identify Risk Groups for Alzheimer's Disease. *J. Alzheimer's Dis.* **57**, 275–283 (2017).
13. Hayden, K. M., Lutz, M. W., Kuchibhatla, M., Germain, C. & Plassman, B. L. Effect of APOE and CD33 on cognitive decline. *PLoS One* **10**, 1–10 (2015).
14. Kauppi, K., Rönnlund, M., Nordin Adolfsson, A., Pudas, S. & Adolfsson, R. Effects of polygenic risk for Alzheimer's disease on rate of cognitive decline in normal aging. *Transl. Psychiatry* **10**, (2020).
15. Caselli, R. J. *et al.* Longitudinal Modeling of Age-Related Memory Decline and the APOE ε4 Effect . *N. Engl. J. Med.* **361**, 255–263 (2009).
16. Harris, S. E. *et al.* Polygenic risk for alzheimer's disease is not associated with cognitive ability or cognitive aging in non-demented older people. *J. Alzheimer's Dis.* **39**, 565–574 (2014).
17. Anne Nelson, E. & Dannefer, D. Aged Heterogeneity: Fact or Fiction? The Fate of Diversity in Gerontological Research. *Gerontologist* **32**, 17–23 (1992).
18. Demidenko, E. *Mixed models: theory and applications with R*. (John Wiley & Sons, 2013).
19. Pinheiro, J. & Bates, D. *Mixed-Effects Models in S and S-PLUS*. (Springer-Verlag, 2000). doi:https://doi.org/10.1007/b98882
20. Zhang, Q. *et al.* Risk prediction of late-onset Alzheimer's disease implies an oligogenic architecture. *Nat. Commun.* **11**, 1–11 (2020).
21. de Rojas, I. *et al.* Common variants in Alzheimer's disease and risk stratification by polygenic risk scores. *Nat. Commun.* **12**, (2021).
22. Escott-Price, V. & Hardy, J. Genome-wide association studies for Alzheimer's disease: Bigger is not always better. *Brain Commun.* **4**, 1–7 (2022).
23. Korologou-Linden, R. *et al.* The causes and consequences of Alzheimer's disease: phenome-wide evidence from Mendelian randomization. *Nat. Commun.* **13**, (2022).
24. Altmann, A. *et al.* A comprehensive analysis of methods for assessing polygenic burden on Alzheimer's disease pathology and risk beyond APOE. *Brain Commun.* **2**, (2020).
25. Simonsohn, U., Simmons, J. P. & Nelson, L. D. Specification curve analysis. *Nat. Hum. Behav.* **4**, 1208–1214 (2020).
26. Steegen, S., Tuerlinckx, F., Gelman, A. & Vanpaemel, W. Increasing Transparency Through a Multiverse Analysis. (2016). doi:10.1177/1745691616658637
27. Knickmeyer, R. C. *et al.* Common Variants in Psychiatric Risk Genes Predict Brain Structure at Birth. 1230–1246 (2014). doi:10.1093/cercor/bhs401
28. Axelrud, L. K. *et al.* Polygenic risk score for Alzheimer's disease: Implications for memory performance and hippocampal volumes in early life. *Am. J. Psychiatry* **175**, 555–563 (2018).
29. Walhovd, K. B. *et al.* Genetic risk for Alzheimer disease predicts hippocampal volume through the human lifespan. *Neurol. Genet.* **6**, (2020).
30. Foo, H. *et al.* Associations between Alzheimer ' s disease polygenic risk scores and hippocampal sub fi eld volumes in 17 , 161 UK Biobank participants. *Neurobiol. Aging* **98**, 108–115 (2021).
31. Chauhan, G. *et al.* Association of Alzheimer's disease GWAS loci with MRI markers of brain aging. *Neurobiol. Aging* **36**, 1765.e7-1765.e16 (2015).
32. Du, J. *et al.* Exploration of Alzheimer ' s Disease MRI Biomarkers Using APOE4 Carrier Status in the UK Biobank. 1–30 (2021).
33. Walhovd, K. B., Lövden, M. & Fjell, A. M. Timing of lifespan influences on brain and cognition. *Trends Cogn. Sci.* 1–15 (2023). doi:10.1016/j.tics.2023.07.001
34. Storsve, A. B. *et al.* Differential Longitudinal Changes in Cortical Thickness, Surface Area and Volume across the Adult Life Span: Regions of Accelerating and Decelerating Change. *J. Neurosci.* **34**, 8488–98 (2014).
35. Salthouse, T. A. Attrition in Longitudinal Data is Primarily Selective with Respect to Level Rather than Rate of Change. *J. Int. Neuropsychol. Soc.* **25**, 618–623 (2019).
36. Ferreira, D., Nordberg, A. & Westman, E. Biological subtypes of Alzheimer disease A systematic review and meta-analysis. **0**, (2020).
37. Roe, J. M. *et al.* Asymmetric thinning of the cerebral cortex across the adult lifespan is accelerated in Alzheimer's Disease - bioxRiv. *bioRxiv* 2020.06.18.158980 (2020). doi:10.1101/2020.06.18.158980
38. Fjell, A. M., McEvoy, L., Holland, D., Dale, A. M. & Walhovd, K. B. Brain changes in older adults at very low risk for Alzheimer's disease. *J. Neurosci.* **33**, 8237–42 (2013).
39. Fjell, A. M. *et al.* One-Year Brain Atrophy Evident in Healthy Aging. **29**, 15223–15231 (2009).
40. Vogel, J. W. *et al.* Four distinct trajectories of tau deposition identified in Alzheimer's disease. *Nat. Med.* **27**, 871–881 (2021).
41. Mohanty, R., Ferreira, D., Nordberg, A. & Westman, E. Associations between different tau - PET patterns and longitudinal atrophy in the Alzheimer ' s disease continuum : biological and methodological perspectives from disease heterogeneity. *Alzheimers. Res. Ther.* 1–16 (2023). doi:10.1186/s13195-023-01173-1

REVIEWER COMMENTS

Reviewer #2 (Remarks to the Author):

2.1. Having had the opportunity to read the revised version of this manuscript, along with the author's detailed responses to my comments, I am happy that most my questions have been resolved. As I understand the main contribution of this study is showing that AD common genetic risk associates with an accelerated general brain ageing, more prominent, though, in the hippocampus. This association not being only driven by the APOEe4, despite this being a strong contributor to it. The methods used, the results presented and the conclusions regarding this appear clearer to me now and I have no major concerns about it.

We thank the Reviewer for their detailed feedback, which has helped improve the paper. We note that another key contribution of our study is the method by which we estimate change. Namely, conditioning change estimates on age both enhances the signal in change estimates, and allows us to use adult lifespan data as test data in machine learning models trained on AD patients. We added the following in the results:

"Across the 576 PRS-AD tests, conditioning brain change estimates on age conferred a 91% strengthening in the observed PRS-AD regression coefficients, compared to using absolute change estimates and correcting for mean age ($t = -9.9$, $p = 2.3e^{-21}$). This was also evident in the subset of results where both change estimates were FDR-corrected significant (54%, $t = -9.01$, $p = 8.1e^{-15}$)."

2.2. I would only suggest rephrasing the author's second hypothesis that, as written, does not read well to me (lines 148-149: "and that variation in AD risk factors would be detectable via brain change rates relative to one's age in AD-sensitive regions in healthy adults") -> which AD risk factors are the authors referring to? And what 'being detectable mean'? I think here they were to suggest that AD PRS will be predictive of a faster decline than expected by age in AD sensitive structures like hippocampus; but not sure it is this by the reading of their statement.

We rephrased the sentence as follows:

"We hypothesized that neurodegeneration in ageing and AD is linked on a continuum, and that individuals changing more than their age would predict in AD-sensitive features across adult life would have quantitatively higher genetic AD risk. We further hypothesized that individuals with both higher genetic risk and faster atrophy for their age would exhibit more memory decline detectable across adulthood"

2.3. The section of the manuscript that I still struggle a bit with, is the analyses involving cognitive function. Admittedly, I do not fully understand the method that the authors applied, but my concern goes beyond that, I don't understand it's purpose and what is its real contribution to the overall study.

We agree we should have clarified the purpose and contribution of the memory change analysis better. Please see our next comment detailing the changes. We also amended the Results to better clarify the four-sample splitting:

“Finally, in the LCBC adult lifespan discovery sample, we separated individuals into discrete groups based on the conjunction of brain and genetic risk factors. We hypothesized that high PRS-AD individuals who are also high on a multivariate marker of relative change in AD-sensitive features would show more pronounced longitudinal memory decline across adult life (pink quadrant 4 in Fig. 6D; Methods).”

2.4. First, I do not see the link between this section and the main results/conclusion; in fact, the authors do not devote much to it either: it is only marginally introduced/justified at the end of the introduction, and it is only discussed briefly with no much emphasis on what brings to the table.

In the Introduction, we make the point that only very weak average effects of genetic AD risk have been found upon cognitive decline in healthy adults, and that this may be due to individual differences in risk also within genetically high-risk groups. This provides the background for our four-sample splitting based on the conjunction of risk markers. We also argue that memory decline is not confined to old age, and thus that normal variation in memory decline rates should be detectable over extended age spans. While brief, we believe we concisely make these points, and thus added only some small clarifiers (below).

We also added a sentence to better clarify the contribution of this analysis in the Discussion (which we disagree was brief; reproduced here):

Introduction
“Finally, several studies suggest that genetic AD risk is subtly related to longitudinal memory decline in healthy older adults[43–45], and one adult lifespan study reported a weak association with decline in cognition on average[46]. Thus, AD risk variants may influence differences in memory decline trajectories that are protracted through life and begin in early adulthood[45–48]. However, the extent to which one’s genetic predisposition influences brain and cognitive outcomes probably differs also between individuals at high genetic risk, which may explain why genetic risk alone is not highly predictive of cognitive change[46,49]. Given that individualized approaches to risk assessment are predicated on assessing the conjunction of risks, considering genetic risk together with an individualized marker of relative brain ageing may improve identification of individuals at higher AD risk in healthy adult lifespan data.”

Discussion
“Individuals at higher genetic risk who also showed more atrophy for their age in AD-sensitive features exhibited more memory decline on average across adult life, compared to genetically at-risk individuals with less atrophy. Hence, knowing one’s genetic risk was insufficient, as it was not necessarily reflected in brain and cognitive outcomes. However, considered together with a multivariate marker of brain change, we found a subset of high PRS-AD individuals whose brain status over time was reflected in a greater drop-off in memory. Thus, our results speak to the importance of considering overlapping risk factors rather than only each in isolation, as we found substantial variation in risk also within genetically high-risk individuals — highlighting that genetic AD risk neither determines nor sufficiently predicts cognitive and brain outcomes. Rather, group differences in memory decline were more driven by brain change differences than by genetic differences, as they persisted when controlling for PRS-AD and *APOE-ε4* but not atrophy. Thus, brain change may be crucial for detecting comparatively at-risk individuals in adult lifespan data. Further, memory decline differences were protracted through adulthood, as they were evident in different age subsets, including comparatively younger adults (e.g., within the age-range 30-65; Supplementary Fig. 17). This indicates neurodegeneration in AD-sensitive regions tracks with memory decline differences that are detectable through adulthood. It also emphasizes that memory decline is a gradual phenomenon not confined to old age. This perspective may be obscured in studies that use clinical tests that do not capture subtle cognitive variations, and estimates of decline relative to a group rather than one’s earlier capacity. Our findings extend previous studies finding PRS-AD[43,44,46] or *APOE-ε4*[45] relates to memory decline across adult life, and possibly shed light

on why such genetic associations are often weak[43–46] or absent[49]. Whether these brain trajectory differences are relevant for later AD outcomes will require follow-up and biomarker assessment, but our results show these neurodegenerative changes are not benign. They also underscore the need for follow-up data over extended age-spans for prediction or prevention of AD, and suggest a continuous view on lifespan brain health may aid understanding of AD[78,79]. Multivariate atrophy measures may help assess AD risk and improve selection into clinical trials. Future research should examine why some high PRS-AD individuals decline more in brain and memory where others remain resilient, as well as combine multivariate change with other biomarkers (e.g., tau, amyloid, inflammation) as we move towards a future of individualized risk assessment."

Which alongside the following sentences, we now feel makes the purpose and contribution of this analysis clear:

"Hence, knowing one's genetic risk was insufficient, as it was not necessarily reflected in brain and cognitive outcomes."

"Thus, tracking change in AD-sensitive regions enhanced the value of knowing a person's genetic risk, and atrophy predicted memory decline more than the genetics"

2.5. Initially I thought this section was aimed at answering the question: 'is it the accelerated brain decline linked to AD genetic risk key in explaining AD cognitive decline, or the speed of decline is secondary to the absolute brain volume?' In other words, does it matter cognitively whether you lose volume faster than expected for your age, or what matter is whether you lose up to certain total amount of grey matter. However, the analyses performed do not allow to test the above.

Our analyses indeed show it matters cognitively if you lose brain volume faster than expected for your age, which we note represents a novel way to quantify brain changes in healthy adults. While interesting, exploring the alternative approach—whether reaching a specific threshold of grey matter loss is more critical—would require either:

- Accurate initial estimates of total grey matter for each individual, to assess brain loss in relative terms. This approach is hampered by variations in subject age at baseline, large differences in brain structural volumes to begin with (Figure 1), and non-linear changes in brain structure that preclude linear estimation of subject-specific volumes at equivalent baseline ages.
- Assuming equivalence across subjects in terms of absolute brain measures (e.g., a 3000 mm^3 hippocampus threshold in everyone) – which would then be similar to a cross-sectional design. But thresholds of grey matter loss are likely to be, at least to a certain extent, individual-specific, and will further depend on factors such as brain and cognitive reserve.

Thus, although an interesting avenue for future research, the proposed approach is hindered by methodological/data issues, and also negates the advantage of our longitudinal design. Our study's strength lies in its ability to estimate how much brain volume individuals lose both relative to their earlier selves, and relative to their age. This approach demonstrates the significance of accelerated brain changes across much of healthy adulthood, as well as their importance for cognitive decline outcomes, among healthy adults tracked across their lives.

2.6. but their four-group breaking of the sample lets to the conclusion that brain changes are potentially the key element predicting cognitive performance (line 505-506) than genetic risk per se; which should not come as a surprise, since one could predict genetic risk to impact cognitive phenotypes via its potential effects on brain??

Although we agree that genetic risk likely impacts cognitive outcomes via its effects on the brain (explaining why genetically exposed individuals with less atrophy showed relatively less memory decline), the message that genetic risk is not necessarily reflected in brain and cognitive outcomes arguably offers an important, nuanced, and somewhat less deterministic view on genetic AD risk. We initially identified the highest risk group as those exhibiting both faster brain change and high genetic risk, and aimed to see if our longitudinal brain data could reveal insights beyond predicting something we already know (i.e., genetic risk). Still, when testing the memory decline difference against all other groups, we were surprised that the clear difference was found between other genetically high-risk individuals. Thus, the primary predictor of memory decline variations in healthy adults turned out to be atrophy in AD-sensitive brain regions, which surpassed the predictive power of both APOE and PRS-AD. The key finding that substantial variation in risk is evident also within high genetic risk individuals is now emphasized throughout the paper:

"Finally, genetically at-risk individuals with more brain change showed more memory decline through adulthood, compared to genetically at-risk individuals with less brain change. Thus, tracking change in AD-sensitive regions enhanced the value of knowing a person's genetic risk, and atrophy predicted memory decline more than the genetics."

Admittedly, it might be me, but I can't really make much sense of what the authors intend to explore here and what the novelty of their results is. Perhaps this should have been spelled more clearly? Allegedly, though, this is a small part of their work that should not take away the merits of the main analyses/results.

We hope we have now clarified the purpose and novelty of this analysis. Finally, we emphasize that our sample consists primarily of high-performing adults (the average IQ of those aged >60 is 119.5, compared with 117.3 in those aged <60), probably due to non-random retention of the most cognitively healthy adults who choose to participate in such studies. We conducted a final analysis to see if this non-random retention was evident in our data, and added a sentence in the Discussion to draw attention to this very important point. Thank you for your review.

Seventh, longitudinal studies inevitably recruit and culminate in unrepresentatively high-performing samples[58]. Our data also suggest this, as we observed a tendency for better memory in older adults with more repeat visits (Supplementary Figure 18), and higher average IQ scores in those older than 60 (Supplementary Table 2). Since even in these we find variation in brain ageing slopes that correlates with AD-related genetic variation and memory decline, the population effect-sizes may be larger.

Reviewer #6 (Remarks to the Author):

Authors have responded to all of the recommendations and have clarified the raised concerns.

Reviewer #6 (Remarks on code availability):

NA
Our code has been available at github.com/jamesmroe/ADchangeRisk since initial submission (see code availability statement).