BMJ Open

Comparisons of neurodegeneration over time between healthy ageing and Alzheimer's disease cohorts via Bayesian inference

Journal:	BMJ Open
Manuscript ID	bmjopen-2016-012174
Article Type:	Research
Date Submitted by the Author:	06-Apr-2016
Complete List of Authors:	Cespedes, Marcela; Queensland University of Technology, Mathematical Sciences School Fripp, Jurgen; CSIRO Digital Productivity and Services, Australian E-Health Research Centre McGree, James; Queensland University of Technology, Mathematical Sciences School Drovandi, Christopher; Queensland University of Technology, Mathematical Sciences School Mengersen, Kerrie; QUT, Statistical Science Doecke, James; CSIRO Digital Productivity and Services, Australian E- Health Research Centre
Primary Subject Heading :	Neurology
Secondary Subject Heading:	Neurology, Epidemiology, Radiology and imaging
Keywords:	Bayesian inference, mixed effects models, neurodegeneration, Alzheimer's disease, longitudinal neuroimaging study

SCHOLARONE[™] Manuscripts

Comparisons of neurodegeneration over time between healthy ageing and Alzheimer's disease cohorts via Bayesian inference

Corresponding author

Marcela Ines Cespedes

Mathematical Sciences School, Queensland University of Technology GPO BOX 2434 Brisbane QLD 4001 Australia Marcela.Cespedes@hdr.qut.edu.au Tel: +61-07-313-87440

Co-authors:

1

3

- Jurgen Fripp CSIRO Digital Productivity and Services Australia E-Health Research Centre Herston, QLD, Australia
- 2 James M. McGree Mathematical Sciences School Queensland University of Technology Brisbane, Australia
 - Christopher C. Drovandi Mathematical Sciences School Queensland University of Technology Brisbane, Australia

- 4 Kerrie Mengersen Mathematical Sciences School Queensland University of Technology Brisbane, Australia
- 5 James Doecke CSIRO Digital Productivity and Services Australia E-Health Research Centre Herston, QLD, Australia

Keywords: Bayesian inference; mixed effects models; neurodegeneration; longitudinal neuroimaging study; Alzheimer's disease

Manuscript word count: 6,000 words

Comparisons of neurodegeneration over time between healthy ageing and Alzheimer's disease cohorts via Bayesian inference

Marcela I. Cespedes · Jurgen Fripp · James M. McGree · Christopher C. Drovandi · Kerrie Mengersen & James D. Doecke

Abstract

Objectives: In recent years, large scale longitudinal neuroimaging studies have improved our understanding of healthy ageing and pathologies including Alzheimer's disease (AD). A particular focus of these studies is group differences and identification of subjects at risk of conversion. For this, statistical analysis using Linear mixed effects (LME) models are used to account for correlated observations from individuals measured over time. A Bayesian framework for LME models in AD is introduced in this paper to provide additional insight often not found in current LME volumetric analyses.

Setting and participants: Longitudinal neuroimaging case study of ageing were analysed in this research on 260 participants diagnosed as either healthy controls (HC), mild cognitive impaired (MCI) or AD. Bayesian LME models for the ventricle and hippocampus regions were used to; (i) estimate how the volumes of these regions change over time by diagnosis, (ii) identify high risk non-AD individuals with AD like degeneration by ranking participants in order of volumetric rate of change, and (iii) determine probabilistic trajectories of diagnosis groups over age.

Results: We observed (i) large differences in average rate of change of volume for the ventricle and hippocampus regions between diagnosis groups, (ii) high risk individuals who had progressed from HC to MCI and displayed similar rates of deterioration as AD counterparts, and (iii) critical time points which indicate where deterioration of regions begin to diverge between the diagnosis groups.

Conclusion: To our knowledge, this is the first application of Bayesian LME models to neuroimaging data which provides inference on a population and individual level in the AD field. The application of a Bayesian LME framework, through simulation methods allows for additional information to be extracted from longitudinal studies. This provides health professionals with valuable information of neurodegeneration stages, and a potential to provide a better understanding of disease pathology.

Strengths of this study

- The models presented in this research easily accommodate for realistic longitudinal neuroimaging settings, which address challenges such as; large patient drop-out (unbalanced design), large and small diagnosis groups and noisy MRI observations.
- Our quantitative analysis and conclusions support those in neuroimaging and neurodegeneration literature.
- This is the first study of its kind to incorporate data external to this analysis, in terms of prevalence rates, in conjunction with the statistical models to infer trajectories over age. Usually this type of inference would require a separate model and analysis with complete and additional data at hand.
- Variability in model estimates are easily visualized and interpreted in terms of credible intervals, box plots and posterior densities. This gives health professionals and people with non-statistical backgrounds various options to assess uncertainty and aid in decision making process.

Limitations of this study

- This research does not accommodate for participants with other neurological disorders or forms of dementia, and assumes participants are in one of three groups; healthy control (HC), mild cognitive impaired (MCI) and Alzheimer's disease (AD).
- Participants with single observations are also assigned a degeneration rank and estimated rate of change, despite not having observed a rate of change in degeneration.
- Additional covariates which are known to affect neurodegeneration were not included in this analysis, such as gender and genetic status.

1 INTRODUCTION

Alzheimer's disease (AD) is the most common form of dementia worldwide ¹. Advances of neuroimaging techniques have been useful for early diagnosis of neurodegenerative disorders ²³ and, coupled with mathematical and statistical models, provide insight to better understand healthy ageing and disease pathology degeneration.^{4–6} The use of linear mixed effects (LME) models has been advocated by Bernal- Rusiel et al. (2013a)⁷ and more recently by Ziegler et al. 2015)⁸ to characterise longitudinal degeneration from neuroimaging Longitudinal neuroimaging Bayesian LME (BLME) models are applied in this research to provide insight into the diagnosis of AD over time. In this research we address three main areas; population diagnosis comparisons based on estimated volumetric rate of change over age, ranking of participants by order of linear volumetric rate of change, and probability trajectories across age of diagnosis groups, conditional on prevalence rates.

Clinical diagnosis classification comparisons are often of interest in longitudinal neuroimaging studies.^{7,9,10} Previous LME models of volumetric degeneration reported on comparisons assessing ranking of diagnosis levels.^{7,11} However in these studies, the magnitude of the differences of disease progression as well as their estimated variances is often excluded,^{7,10–13} thus a richer insight into the differences of diagnosis levels is lacking. The BLME approach uses simulation techniques to draw from the posterior distribution, which is a combination of prior information and information from the data (through the likelihood function), to assess diagnosis group estimates and comparisons. These simulations quantify uncertainty and provide posterior probabilities that can be compared directly, without referring to significance levels or multiple statistical tests.

The development of methods which account for large inter and intra variability of biomarkers presents a challenge in longitudinal neuroimaging studies.^{14–16} Furthermore, the observations of diagnosis group tends to become unbalanced over time which makes it difficult to deduce information of the complex AD pathway. However, insight into neurodegeneration of high risk participants, namely mild cognitive impairment converters, is crucial for early detection methods and improving diagnostic accuracy of AD.^{5,17} Several authors such as Harville and Carriquiry (1992),¹⁸ Gelman and Hill (2006)¹⁹ and Li et al. (2012)²⁰ state BLME models have the capability to seamlessly handle unbalanced data and small-sample de- sign analysis. This motivates our choice of statistical framework, as we aim to utilise as much information as possible from the study analysed, and retain participants with a single observation.

Individuals ranked by order of neurodegeneration severity allows for comparisons of progression of all individuals over the study time, while quantifying the uncertainty and estimating variability of individualised conversion rates. The application of BLME models allows for estimation of class membership probabilities and estimation of deterioration rates of each participant via analysis of random effects. This type of analysis is often overlooked in longitudinal studies of ageing.²¹

As the field of neuroimaging in AD has been rapidly expanding in the past 20 years,^{14,22–24} it is of interest to incorporate as much relevant information as possible, as independent longitudinal neuroimaging studies often build on and support each other.^{5,25,26} This can be achieved using the Bayesian approach, as it combines external information with experimental data at hand, while accounting for various sources of uncertainty. Often this background information can be incorporated in the form of the prior, but it can also be applied after estimation of the model to provide additional inference from our model outcomes. In the current project, we demonstrate this concept by combining model information with prior knowledge obtained from prevalence studies to formulate probabilistic diagnosis group trajectories over age.

Authors Jack et al. (2014)²⁷ highlight the importance of population frequency or probabilistic trajectories of neurodegeneration groups over a wide age span. Their study quantified frequencies of expected neurodegeneration cases de- pendent on ages 50 to 89. Particular focus was placed on asymptomatic individuals (preclinical AD) who were at risk of developing AD and ages of increased frequency of convergence to AD as they reach their later years. While our methods can also be used for similar purposes and place emphasis on a particular neurodegeneration group, the goal for our final analysis is to identify critical time points where all diagnosis levels begin to diverge. This can aid in discovering groups or patterns in neurodegeneration consistent with healthy ageing or the AD pathway. Alternatively a similar analysis can also be used to compare diagnosis trajectories of different longitudinal neuroimgaing population studies, such as the Alzheimer's Disease Neuroimaging Initiative (ADNI).

This paper is outlined as follows. Section 2 describes the case study, Sections 3.3 and 4.1 show an application of the BLME models to address multiple comparisons of various sizes from baseline diagnosis, including large For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

BMJ Open

BMJ Open: first published as 10.1136/bmjopen-2016-012174 on 7 February 2017. Downloaded from http://bmjopen.bmj.com/ on June 13, 2025 at Agence Bibliographique de Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

(healthy control, $N_{HC} = 168$ people) and small groups (mild cognitive impaired, $N_{MCI} = 50$ people and Alzheimer's disease, $N_{AD} = 42$ people at baseline). Sections 3.4 and 4.2 rank individuals by order of neurodegeneration severity, thereby comparing the progression of all individuals over the study time. Approximately 10% of individuals convert from baseline case to a worse diagnosis throughout the length of the study. This analysis allows for the identification of those participants who are most at risk of developing AD like rates of deterioration for the hippocampus and ventricle regions of the brain. The third and final area addressed in this research is presented in Sections 3.5 and 4.3, which estimates probabilistic diagnosis group trajectories across age, derived from neuroimaging information. This requires the synthesis of in- formation from the study cohort and the AD literature.

2 AIBL LONGITUDINAL STUDY OF AGEING

The neuroimaging data analysed in this paper were obtained from the Australian Imaging Biomarker and Lifestyle Study of Ageing (AIBL). This is an ongoing study which aims to discover which biomarkers such as cognitive assessment results, neuroimaging, lifestyle and demographic factors potentially influence subsequent development of AD. The sample comprises N = 260 people, who have at most four repeated observations approximately 18 months apart. These data are highly unbalanced, since patient drop out occurs at every time point throughout the study, with approximately 69% of participants in the final follow up.

Key regions of the brain which are strongly associated with neurodegeneration in relation to AD and healthy ageing include the lateral ventricles^{28,29} and hippocampus volumes.^{3,30–33} Atrophy due to disease pathology spreads throughout particular regions such as the hippocampus, which leads to a general decrease in volume over time. The decrease in brain matter results in an increase of cerebrospinal fluid (CSF) which bathes and cushions the brain and spinal cord. The lateral ventricles are filled with CSF, hence an increase of overall brain atrophy results in an increase of ventricle volume. Models presented here were considered separately for the lateral ventricles and the sum of the left and right hippocampal (hippocampus) volume derived from MRI. See Rowe et al. (2010)³⁴ for details on image acquisition and processing.

Brain region volumes were normalised by the intracranial volume (ICV), hence all volumes are in the (0, 1) interval. This accounts for the variability of different cranial sizes, while preserving the trend in volume.^{35,36} Due to the wide range in values and in order to eliminate numerical problems in the estimation of these models, age was standardised $(age - \overline{age})/sd(age)$, where age and sd(age) are the empirical mean and standard deviations of the study group ages. Likewise, the hippocampus ICV response was scaled up by a factor of 100, in order to avoid variance estimates close to zero which can be difficult to estimate. All participants in this study were categorised as: healthy control (HC), mild cognitive impaired (MCI) and those with a probable diagnosis of Alzheimer's disease (AD) at each time point based on neuropsychological diagnosis. The aim of these models was to capture the linear decrease in regional brain volume across ages for people within three diagnosis groups.

3 METHODS

Linear Mixed Effects (LME) models are a standard approach to modelling repeated observations from several individuals.³⁷ Standard LME models re- quire the following assumptions to be met: a linear relationship exists between the response and the explanatory variables; the response and error terms at every level are Gaussian although for non-normal models we may extend this assumption to the exponential family and apply generalised linear mixed models;³⁸ the variances across all levels are homoscedastic, and repeated observations for an individual can be correlated, but observations between people are assumed independent. The general LME model is of the following form,

$$y = X\beta + Z\boldsymbol{b} + \varepsilon$$

(1)

where X and Z denote design matrices, and vectors β and **b** are the fixed and random effects respectively for r fixed and m random effects. The residual vector ε is assumed to be normally distributed with $\varepsilon \sim N(0, \sigma^2 I_r)$, where I_r is the $r \times r$ identity matrix. The random effects vector **b** is assumed to be multivariate normally distributed, $\mathbf{b} \sim MVN(\mathbf{0}, \Sigma)$, where the variance covariance matrix of the random effects is denoted by Σ .

3.1 Statistical analysis

In a Bayesian framework the likelihood corresponding to the model in equation (1) is $p(y|\beta, X, Z, b, \sigma^2, \Sigma)$, which is conditional on the random effects and on the model parameters. The resultant joint posterior distribution for the model parameters and random effects given the data is

$$p(\beta, \boldsymbol{b}, \sigma^2, \boldsymbol{\Sigma} | \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{y}) \propto p(\boldsymbol{y} | \beta, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{b}, \sigma^2, \boldsymbol{\Sigma}) \ p(\boldsymbol{b}, \boldsymbol{\Sigma}) p(\beta) p(\sigma^2) p(\boldsymbol{\Sigma}).$$
⁽²⁾

In the absence of external information, weakly informative priors, $p(\beta)p(\sigma^2)$ and $p(\Sigma)$ were used throughout; refer to expression (3) in Section 3.2 for full specification of priors. Under the Bayesian paradigm all the assumptions stated in Section 3 remain. Furthermore as Gelman et al. (2013)³⁹ and Gelman and Hill (2006)¹⁹ state, additional complexity and generalisation of the LME model comes naturally under the Bayesian framework.

Estimation of the model parameters was achieved by sampling from the joint posterior distribution using Markov chain Monte Carlo (MCMC) techniques⁴⁰ which samples from the marginal posterior distributions as a by-product. Note that the parameter estimates are obtained by integrating over the posterior distribution, rather than maximising the likelihood, as numerical methods to solve for integrals in high dimensions are often difficult to compute.^{40,41}

3.2 BLME in the context of the case study

 Following equation (1), the normalised volume is denoted by Y_{ij} for the i^{th} individual at the j^{th} time point, where binary values x_{MCI} and x_{AD} refer to the two levels of diagnosis, MCI and AD respectively with HC as the baseline. The Bayesian LME model for person i = 1, 2, ..., 260 at time point j = 1, 2, 3, 4 is given by

 $Y_{ij} \mid \mu_{ij}, \sigma^{2} \sim N(\mu_{ij}, \sigma^{2})$ $\mu_{ij} = \beta_{0} + \beta_{1} x_{MCI,ij} + \beta_{2} x_{AD,ij} + \beta_{3i} StndAge_{ij} + \beta_{4i} StndAge_{ij} x_{MCI,ij}$ $+ \beta_{5i} StndAge_{ij} x_{AD,ij}$ $\beta_{ki} = \beta_{k} + b_{ki}, for k = 0, 3, 4, 5$ $\boldsymbol{b}_{i} \sim MVN(\boldsymbol{0}, \Sigma).$ (3)

Random effects $\mathbf{b}_i = [b_{0i}, b_{3i}, b_{4i}, b_{5i}]$ denotes the i^{th} individuals deviation from population means $\beta_0, \beta_3, \beta_4$ and β_5 . The model in (333) allows for correlation between random effects and this is reflected by the structure of the priors. The residual variance and the variance-covariance matrices are designated by semi-conjugate priors $\sigma^2 \sim IG(0.001, 0.001)$ and $\Sigma \sim Wishart(R, 6)$ respectively, where $= 1,000 \times I_4$. The fixed effects vector $\boldsymbol{\beta} = [\beta_0, \beta_2, \beta_3, \beta_4, \beta_5]^T$ is assumed normal with $\boldsymbol{\beta} \sim MVN(\mathbf{0}, 1e6, \times I_6)$. Posterior predictive plots were used as a measure of goodness-of-fit. This involved simulating from the posterior distribution and forming 95% credible intervals of the posterior predictive responses, which were compared with the observed responses.

The R software was used to implement the Bayesian models.⁴² The rjags package⁴³ implemented MCMC methods to estimate the parameters. Packages coda⁴³ and ggplot2⁴⁴ were used to analyse the MCMC chains and visualise the three sets of analyses presented here. All R source code for this manuscript and simulated data is available at github website <u>https://github.com/MarcelaCespedes/Bayesian inference on neuroimaging</u>.

Two independent MCMC runs were performed using different starting values; each chain ran for 300K iterations of which 100K were discarded as burn-in and the remaining simulations were thinned at every 50th iteration. The retained 8,000 simulations were used to estimate the posterior distributions. Convergence diagnostics of the chains included observing the trace, density and autocorrelation plots as well as the Gelman and Rubin diagnostic.⁴⁵ Desirable chain mixing and convergence was observed in all diagnostics. For posterior checks and diagnostics refer to the supplementary material.

3.3 How do HC, MCI and AD participants degenerate over time?

Performing a Bayesian analysis provides a posterior distribution of the parameter which here can be used to estimate the rate of volumetric degeneration for each diagnosis level.¹² In this analysis, we estimate a diagnosis group effect via the posterior mean of the relevant parameter, and investigate differences in these effects via credible intervals (about differences of these means). Other than mean diagnosis comparisons, further analysis in terms of mean differences of these groups is often not performed in LME volumetric neuroimaging models.^{7,46} However as highlighted in Apostolova et al. (2012)²⁸ and Holland et al. (2012)³², such insight allows for potential techniques to detect signs of AD like neurodegeneration on pre-symptomatic individuals.

BMJ Open

As indicated in (3), the population rate of deterioration for each diagnosis consists of the addition of the baseline effect (HC) with the interaction terms for the other diagnosis groups (MCI or AD). Thus the posterior marginal distributions of β_3 for the baseline, $\beta_3 + \beta_4$, $\beta_3 + \beta_5$ for MCI and AD diagnosis respectively, were compared.

Furthermore, the order of deterioration of the diagnosis levels over both brain regions was assessed. Posterior probabilities were used to order parameter values, since this allows for direct probabilistic diagnosis group comparisons based on the MCMC output while quantifying uncertainty in the parameter estimates. Let M be the number of MCMC posterior draws; in our methods M = 8,000 as described in Section 3.2.

The probability that the rate of change for MCI degeneration is smaller than an AD diagnosis for the ventricle region is estimated by

$$P(MCI < AD) = \frac{1}{M} \sum_{m=1}^{M} \mathbb{1}(\beta_3^m - \beta_5^m < 0)$$
⁽⁴⁾

Where the indicator function 1 is equal to 1 if $\beta_3^m - \beta_5^m$ and 0 otherwise. Probabilities for other comparisons of diagnosis levels for the ventricle and hippocampus regions are computed in a similar manner; see Section 4.1 for full results.

3.4 How to identify individuals with high levels of neurodegeneration?

It is expected that individuals who are healthy (HC) will have relatively minimal deterioration while those with MCI or AD will show increasing levels of deterioration. Hence we would expect that the volumetric rate of change will reflect the neuropsychological clinical diagnosis. However as noted by Woolrich et al. (2004);⁴⁷ Bernal-Rusiel et al. (2013a)⁷ and Bernal-Rusiel et al. (2013b)⁴⁸, high inter and intra variability is often observed in longitudinal neuroimaging studies. For this reason, in this analysis we foresee the estimated volumetric rate of change for a few individuals not to group with participants of the same diagnosis and exercise caution when comparing estimated trajectories of individuals with a single observation.

Participants with outlier rates of deterioration or not within range of their diagnosis levels, as well as those who con-verted throughout the study are of particular interest as they do not conform to the larger gradient over time ordering. Thus a question of interest might be: if an individual has a high neurodegeneration rate with respect to their corresponding diagnosis group, are they likely to degenerate along the AD pathway?

In the imaging sub-cohort of the AIBL study, three individuals progressed directly from HC to AD, eight people progressed from HC to AD and a further 16 individuals progressed from MCI to AD. These converters can be tracked to observe their severity with respect to the rest of the cohort. In this section, particular focus is on the converters who progressed from HC to MCI, and the comparison of their estimated rates of deterioration with AD participants, as they could be potential AD converters of neurodegeneration, to estimate their probability of remaining in such high rank.

Unlike our first analysis, which compared the estimated population effect across all diagnosis levels, the focus here is on an individual's rate of deterioration. The marginal posterior distributions of individual random effects values in the HC (β_{3i}), MCI ($\beta_{3i} + \beta_{4i}$) and AD ($\beta_{3i} + \beta_{5i}$) groups are inspected, to estimate the rate of deterioration, for i = 1, 2, ..., 260 individuals on all four time points.

Furthermore, as discussed in Section 3.4, the ordered box plots in Fig. 4 illustrate median rankings of participants and illustrate the large variation between individuals. Distribution of ranks on participants take into account the high variation between individuals, by ranking participants at every iteration of the MCMC simulation of the random effects. This results in M = 8, 000 simulations on every individual and allow us to derive probabilistic statements on individuals of interest remaining in a specified ranking range, for example the top 15^{th} quantile. This analysis was performed on both a subset of the data, using observations with the first three time points as well as on the full data (four time points) to investigate the change of rank probabilities over time for particular individuals of interest. Such analysis extends the BLME models to allow the identification of high risk converters among the participants analysed. Full results are described in Section 4.2.

3.5 How do diagnosis trajectories vary over age?

The ventricle and hippocampus models derived in (3) were used to compute probabilities $P(\tilde{y}|HC, age)$, $P(\tilde{y}|MCI, age)$ and $P(\tilde{y}|AD, age)$, for a specified age with volume range denoted by \tilde{y} . Given information available

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

(6)

on an individual at an early age and within the limits of our data age span, we seek to answer: at this early age, for a given volume range (which can be constrained or unbounded), what is the probability that this new individual will be diagnosed as HC, MCI or AD? Moreover, how does this change as the individual ages? These probabilities are estimated below.

At a given age for both ventricle and hippocampus models stated in (3) with diagnosis levels $Diagnosis = \{HC, MCI, AD\}$, the following holds

$$P(AD|\tilde{y}, age) = \frac{P(\tilde{y}|AD, age)P(AD|age)}{\sum_{d=1}^{3} P(\tilde{y}|Diagnosis_{d}, age)P(Diagnosis_{d}|age)}$$
(5)

The BLME model estimates $P(\tilde{y}|AD, age)$, $P(\tilde{y}|MCI, age)$ and $P(\tilde{y}|HC, age)$. As M is the number of MCMC posterior draws, then

$$P(\tilde{y}|AD, age) = \frac{1}{M} \sum_{m=1}^{M} \mathbb{1}(\tilde{y}_m \in \tilde{y}),$$

where the indicator function 1 is equal to 1 if $\tilde{y}_m \in \tilde{y}$ and 0 otherwise. This expression is the average number of predicted values \tilde{y}_m which fall within \tilde{y} . A similar expression was used for MCI and HC diagnosis levels. Probabilities P(HC|age), P(MCI|age) and P(AD|age) were obtained from Ward et al. $(2012)^{49}$ and Refshauge and Kalisch $(2012)^{50}$ for ages 60, 65, 70, 75, 80 and 85. We acknowledged that these are very broad estimates which are generalised over genders, genetic status and many other factors which are known to affect prevalence rates. These prevalence rates also do not take into account participants who develop other forms of dementia, or any other neuropsychological disorders. Refer to the Supplementary material for the full table of probabilities used in this analysis. Similar computations were performed for the other diagnosis levels, MCI and HC, to evaluate related probabilities. Due to the wide variability observed in the hippocampus and ventricle volumes among participants, the volume regions were divided into four different ranges, \tilde{y} , which vary over age groups. Quantile growth curves discussed in Cole and Green (1992)⁵¹, and Koenker (2005)⁵², highlight the advantages of algorithms that can estimate non-crossing quantiles which are monotone increasing over age to reflect the heteroscedasticity often found in biological systems. In this paper we utilised the algorithm discussed in Muggeo et al. (2013)⁵³, as it addresses all of these issues and is available via R package quantregGrowth. The \tilde{y} values of took on ranges; $75 - 100^{th}$, $50 - 75^{th}$, $25 - 50^{th}$ and $15 - 25^{th}$ percentile of observed response values, as shown in Fig. 1.

Fig. 1: Percentile ranges of volume across ages 60 to 85 years old, for ventricle (left) and hippocampus (right). Recall region volumes are normalised by the ICV value as they represent a percentage of volume within the intracranial cavity. Ranges up to the 100^{th} percentile henceforth denote the empirical maximum volume for that region. Volume percentiles; $75 - 100^{th}$ from blue (0.75) to top dotted line, $50 - 75^{th}$ from green (0.25) to blue (0.75) line, $25 - 50^{th}$ from red (0.25) to green (0.50) line and $15 - 25^{th}$ from black (0.15) to red (0.25) line.

For completeness in our analysis, volume ranges such as $5 - 25^{th}$ percentile were explored. However there was very little difference in the probability trajectories among these volume ranges, hence we maintained the $15 - 25^{th}$ percentile range. Furthermore, we wished to avoid low volume outliers, and place emphasis on the degenerating trends present in the majority of the data, for biologically meaningful inferences.

The results from applying Eq. (5) show probability trajectories of an individual being in one of the three diagnosis levels, across ages 60-85 within the four quantile ranges. The goal for this analysis is to identify critical time points where diagnosis levels begin to diverge which can aid in discovering groups or patterns in neurodegeneration consistent with healthy ageing or the AD pathway. Furthermore, the influence of covariates gender and apolipoprotein-E (APOE) was explored, by repeating this analysis on sub-groups of male, female, APOE positive and negative.

4 RESULTS

4.1 How do HC, MCI and AD participants degenerate over time?

The atrophy patterns for the ventricle and hippocampus regions described in Section 2 are reflected in the results of the BLME models. A decrease of hippocampus volume and an increase of ventricle volume is depicted by the posterior densities for the rates of deterioration for the two responses as shown in Fig. 2. As expected, this

BMJ Open

biological pattern across the three levels of diagnosis is reflected in Fig. 2 as well as in Tables 1 and 2. The ventricle population estimates of deterioration show an increase of volume as the diagnosis progressively worsens and the hippocampus population estimates of deterioration reflect a decreasing negative slope from HC, to MCI and AD. The overlapping densities are expected as individuals generally progress gradually in order of deterioration, from HC to MCI to AD. Despite this overlap, there are distinct differences between the average rate of volumetric deterioration between the three diagnoses, as seen in Table 1.

Fig. 2: Posterior densities of population mean estimates of linear deterioration rate for diagnosis (top plot): HC, MCI and AD, for ventricle (left) and hippocampus volume (right) models. Dotted lines on bottom plots denote the means for each density, whose values are shown in Table 1.

Tables 1 and 2 present estimated rates of change as well as the probabilities of diagnosis ordering for the hippocampus and ventricles. Furthermore, the difference among HC and progressed degeneration levels MCI and AD, show the additional annual standardised age rate of change which make these groups differences. The increasing credible intervals for each group as deterioration progresses from HC to MCI to AD, illustrate the stratified structure of different sample sizes over groups in our data. The box plots in Fig. 4 also demonstrates the general variability due to various diagnosis sample numbers

Table 1: Top: Posterior means for rates of deterioration across three diagnosis levels for ventricle and hippocampus volume, credible intervals for estimates in parenthesis. Bottom: Group differences among the three diagnosis levels.

		Regions: units IC	/ volume/StndAge	
	Parameter	Ventricle	Hippocampus	
HC	β3	5.6e-3 (4.3e-4 , 6.3e-3)	-1.3e-2 (-1.8e-3 , -9.6e-3)	
MCI	β3 + β4	6.6e-3 (4.6e-3 , 8.8e-3)	-2.2e-2 (-3.1e-2,-1.4e-2)	
AD	β3 + β5	9.6e-3 (5.7e-3, 1.2e-3)	-2.8e-2 (-4.3e-2 , -1.5e-2)	
	Estimated difference of volumetric change among diagnosis groups			
HC - MCI	β4	9.4e-4 (-6.7e-4 , 2.6e-3)	-8.6e-3 (-1.8e-2 , -7.8e-4)	
HC - AD	β5	3.8e–3 (7.9e–4 , <mark>7.0e–</mark> 3)	-1.5e-2 (-3.0e-2 , -1.6e-3)	

Table 2: Posterior probabilities showing comparisons between HC (healthy control), MCI (mild cognitive impaired) and AD (Alzheimer's disease) for ventricle and hippocampus volume. These results provide strong evidence regarding the order of diagnosis levels, derived from the case study.

8 8 8		
Ventricle	P(HC <mci) <="" =="" p(0="" td="" β4)<=""><td>P(MCI<ad))<="" <="" =="" p(β4="" td="" β5=""></ad)></td></mci)>	P(MCI <ad))<="" <="" =="" p(β4="" td="" β5=""></ad)>
	0.980	0.991
Hippocampus	P(AD <mci) <="" =="" p(β5="" td="" β4)<=""><td>P(MCI<hc) 0)<="" <="" =="" p(β4="" td=""></hc)></td></mci)>	P(MCI <hc) 0)<="" <="" =="" p(β4="" td=""></hc)>
	0.806	0.985

Our BLME models also allow for probability statements to be made, based on whether any of the slopes are greater or smaller than a biologically meaningful constant or threshold. Table 2 shows the posterior probabilities of deterioration ordering for the three diagnosis categories for ventricle and hippocampus volume, as computed in Eq. (4).4444 The large probabilities support the sequential pattern of deterioration for both regions.

4.2 How to identify individuals with high levels of neurodegeneration?

The rate of deterioration (as measured by the rate of change with respect to age) for the ventricle and hippocampus are in reverse order; large positive ventricle slopes denote high atrophy whereas low negative slope denote large hippocampus atrophy. Table 3 shows a snippet of the participants ranked in order of their estimated median posterior deterioration rate. The data available in this study are highly unbalanced; nonetheless all individuals are ranked despite 19 patients being observed at a single time point only. This is due to the "borrowing strength" aspect of mixed effects models, in that information across all time points contribute to the estimation of the population trends.

Fig. 4 shows clusterings based on HC, MCI and AD participants, denoted by the blue, purple and red box, respectively. This reflects the general order of diagnosis rates of deterioration for the ventricle and hippocampus For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

volumes as shown in Fig. 2. However, there are a few individuals who do not follow this pattern, namely those in the small clustered group with the positive estimated levels of atrophy in the hippocampus model and participant ID 1122 in the ventricle model. Participant ID's 1122 and 483 are two out of the 19 individuals who only had baseline measurements, so the rate of deterioration was not observed, but it was still estimated.

Table 3: Ranking of individuals from largest to smallest in order of posterior expected rate of deterioration $(\beta_{3i}, \beta_{3i} + \beta_{4i}, \beta_{3i} + \beta_{5i})$ slope for all 260 participants, with 95% credible intervals in parenthesis. Snippet of table shows first and last five individuals, for ventricle and hippocampus volumes. Diagnosis levels; HC, MCI, AD and Converter (either from HC to MCI, HC to AD or MCI to AD), to identify the 27 individuals who changed diagnosis throughout the study, as seen in Fig. 4.

	Ranking	ng AIBL.ID Diagnosis Posterior mean rate of deterioration for indiv		Posterior mean rate of deterioration for individuals
				(credible intervals)
Ventricle	1	1122	AD	-1.4e-3 (-1.1e-2, 9.8e-3)
	2	68	HC	2.3e-3 (-1.9e-3, 6.4e-3)
	3	771	HC	2.8e-3 (-1.3e-3, 6.6e-3)
	4	814	HC	2.9e-3 (-5.4e-4, 6.4e-3)
	5	698	HC	2.9e-3 (-1.1e-4, 6.0e-3)
	256	1032	AD	1.6e-2 (5.5e-3, 2.7e-2)
	257	102	AD	1.6e-2 (8.1e-3, 2.4e-2)
	258	10	AD	1.7e-2 (7.8e-3, 2.7e-2)
	259	658	AD	1.7e-2 (9.4e-3, 2.5e-3)
	260	1102	AD	2.3e-2 (1.5e-2, 3.2e-2)
Hippocampus	1	10	AD	-9.3e-2 (-1.6e-1, -3.7e-2)
	2	12	AD	-6.1e-2 (-9.9e-2, -2.4e-2)
	3	1135	AD	-5.7e-2 (-1.1e-2, -1.0e-2)
	4	398	AD	-5.6e-2 (-9.5e-2, -1.7e-2)
	5	19	AD	-5.4e-2 (-9.8e-2, -1.7e-2)
	256	156	HC	2.0e-4 (-1.5e-2, 1.8e-2)
	257	62	HC	4.1e-3 (-9.5e-3, 1.8e-2)
	258	80	HC	1.0e-2 (-5.4e-3, 2.7e-2)
	259	483	MCI	1.3e-2 (-1.9e-2, 4.6e-2)
	260	1122	AD	1.5e-2 (-1.5e-2, 4.9e-2)

There were 27 individuals who progressed from the baseline case to a worse diagnosis. Eight individuals of interest are those who progressed from HC to MCI and who had at least three repeated observations recorded. Their estimated deterioration rankings are shown in Fig. 4. The majority of the eight converters in the hippocampus model are scattered along the lower half of the ranking of deterioration. This suggests that their linear rates of hippocampus neurodegeneration are less than those of the AD patients. However patient IDs 757, 232 and 471 were ranked approximately mid-way in this analysis, suggesting that they are approaching hippocampus rates of deterioration similar to AD, and out of the eight converters, they are the most at risk.

Likewise for the ventricle model at the top of Fig. 4, patient ID 471 shows a ventricle rate of deterioration strongly similar to the AD cohort. Further investigation of patient ID 471, such as past family mental history of other forms of dementia, stroke or other mental illness, current cognitive status and other health related factors may provide further insight as to why this individual has an unusually high rate of ventricle deterioration in comparison with the rest of the HC to MCI converters.

Fig. 3: Posterior distribution of ranks for converters ID 721, 365 and 12, for Ventricle (top) and hippocampus (bottom) ICV volume models. These density rankings were derived with observations from time points one to three.

Fig. 3 shows the posterior distribution of ranks for participants IDs 721 and 12 who converted from MCI to AD at time point 4. Probabilities of these individuals ranked in the lowest 15^{th} quantile for the ventricle volume is 0.75 and 0.46 respectively for participants IDs 721 and 12, likewise for the hippocampus region, these probabilities For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

BMJ Open

are 0.47 and 0.58. This same analysis can be performed on any quantile range for any participants of interest. These probabilities show that these participants are in the high neurodegeneration extreme. This same analyses on the full data (over four time points) result in probabilities of participants IDs 721 and 12 ranked in the top 15^{th} quantile are 0.80 and 0.66 for the ventricle and 0.54 and 0.69 for the hippocampus regions. Refer to Supplementary material for posterior ranks distribution plots for all 27 converters.

Fig. 4: Box plots of posterior distribution of random effect values for participants in AIBL study (N = 260) for full data (four time points). Ventricle (top) and hippocampus (bottom) rates of deterioration for each participant in the study. As there are 157 HC, 34 MCI, 42 AD and 27 Converters in this study, there is higher uncertainty on the rate of deterioration of converters, MCI and AD participants (hence longer box plots) as compared to the HC (narrower box plots). Eight individuals who converted from HC to MCI throughout the study are highlighted in red with corresponding ID numbers.

4.3 How do diagnosis trajectories vary over age?

The aim of these analyses is to show the relationship between a volume percentile, combined with results from external sources, to predict diagnosis changes over time. As described in Section 3.5, here we present the probability of a new individual diagnosed as either HC, MCI or AD conditional on volume range and specified age between 60-85 years.

Volume ranges, \tilde{y} , were the 75 – 100th, 50 – 75th, 25 – 50th and **15** – **25th** percentiles, as shown in Fig. 1 in Section 3.5. Eq. (5) established relationships $P(HC|\tilde{y}, age)$, $P(MCI|\tilde{y}, age)$ and $P(AD|\tilde{y}, age)$ which consists of the output from BLME model stated in (3) in conjunction with prevalence rates from Ward et al. (2012)⁴⁹ and Refshauge and Kalisch (2012)⁵⁰.

Fig. 5: Probability curves show the posterior probability of HC, MCI or AD diagnosis for the ventricle (top) and hippocampus (bottom) models, 95% interval denotes Monte Carlo error based on several simulations of the BLME models. Total volume divided into four percentile volume ranges, as shown in Fig. 1. Percentiles; $75 - 100^{th}$, $50 - 75^{th}$, $25 - 50^{th}$ and $15 - 25^{th}$.

Uncertainty in the convergence trajectories of diagnosis levels is presented in terms of probabilities, hence no credible intervals can be estimated. However, there is Monte Carlo error associated with these estimates as they are de- rived from a finite sample from the posterior distribution. Due to simulation running time, the ventricle and hippocampus models in expression (3) were estimated independently B = 10 times, hence every computation to derive the probability trajectories in this analysis was also estimated ten times in order to compute the Monte Carlo standard error estimates. Let the estimated quantity be denoted as θ and sd be the standard deviation, then a 95% interval for the Monte Carlo standard error is estimated as $\hat{p} \pm 1.96 \times sd(\theta_1, \theta_2, ..., \theta_B)\sqrt{B}$. As $B \to \infty$, the Monte Carlo standard error tends to zero, and while practically B must be finite, our narrow confidence intervals in Fig. 3.5 suggests our simulation methods are adequate for our application.

The results in Fig. 5 show a large difference between HC in contrast with MCI and AD diagnosis for ages 60 to 75 across all ventricle volume quantiles. From age 75 onwards, those individuals in the top percentile range $(75 - 100^{th})$ show the quickest convergence of all the diagnosis levels, who by age 85, show an approximate equal probability (0.30 and 0.31) of being diagnosed as MCI or AD and only a slightly higher chance (0.39) of remaining HC. This contrasts those participants in the lower ventricle volume range $(15 - 25^{th})$, whose difference in diagnosis is vastly different towards the later ages. By age 85, there is a mean estimated 0.60 probability of remaining HC, 0.27 probability of being classified as MCI and an approximate 0.13 probability of AD diagnosis.

The hippocampus model results for this analysis are shown on the bottom of Fig. 5. Between the ages 60-70 there is very little difference across the diagnosis patterns, suggesting individuals whose hippocampus volume lie above the 15^{th} percentile have an approximately equal risk of HC, MCI or AD diagnosis. From age 70 onwards a noticeable difference in diagnosis trajectories is seen across all volume regions, 5 years earlier than the ventricle volume results. This is supported by a large body of literature^{5,23,28,54–56} as the hippocampus is affected at early stage of development of AD compared to other brain regions. As low hippocampus volume denotes high atrophy, individuals who fall in the lower range volumes, $15 - 25^{th}$ percentile, are most at risk of proceeding onto AD.

Individuals in the lower hippocampus volume range, at age 85 have an approximate equal chance of HC, MCI or AD, as shown in Fig. 5.

Diagnosis trajectories over groups; male, female, apolipoprotein $\varepsilon 4$ (APOE $\varepsilon 4$) carriers and non-carriers were also investigated for the hippocampus and ventricle regions utilising model (3). We assumed the same prevalence rates within the population, for example P(MCI|age) = P(MCI|age, female), hence the same broad prevalence rates from Ward et al. (2012)⁴⁹ and Refshauge and Kalisch (2012)⁵⁰ were used. Very little difference in probable disease trajectory across all groups between ages 60 to 85 were observed (refer to the Supplementary material for plots). APOE $\varepsilon 4$ has been associated to an increased likelihood of developing AD.^{57–59} Gender differences regarding the prevalence of AD have also been studied.^{60,61} As the BLME models and inference derivation presented in this paper are the first of their kind, the objective of this analysis is to demonstrate probable diagnosis trajectories conditional on very broad, non-group specific prevalence rates. Future models which account for APOE- $\varepsilon 4$, gender and other factors will utilise group-specific prevalence rates. However to derive the same inference, this would require group specific prevalence rates across ages 60-85, which are difficult to attain from literature.

Fig. 5: Probability curves show the posterior probability of HC, MCI or AD diagnosis for the ventricle (top) and hippocampus (bottom) models, 95% interval denotes Monte Carlo error based on several simulations of the BLME models. Total volume divided into four percentile volume ranges, as shown in Fig. 1. Percentiles; $75 - 100^{th}$, $50 - 75^{th}$, $25 - 50^{th}$ and $15 - 25^{th}$.

Our results support those presented in Holland et al. (2012)³², whereby diagnosis trajectories for neurodegenerated individuals (that is those with very low hippocampus and high ventricle volume) converge at the highest age group, in general over the age of 85. In particular, our results support those presented in Jack et al. (2014)²⁷ for probabilistic trajectory of beta amyloid negative and neurodegeneration positive participants. To make our results comparable to those from Jack et al. (2014)²⁷, HC participants whose hippocampus volume is less than the 50th percentile are defined as neurodegeneration positive. While both methods present trajectories for neurogedegenration of participants over age, the BLME models presented here primarily estimate the rate of volumetric change for the ventricle and hippocampus regions. There are many other inferences that can be deduced from a combination of tapping into the vast wealth of AD research,^{5,62} coupled with the present study analysis. The results presented here are some of the advantages of modelling neurodegeneration through mixed effects models in the Bayesian framework.

5 Discussion

In this research we extended the level of insight commonly derived by LME models applied to longitudinal neuroimaging data into three key areas based on a BLME model on the ventricle and hippocampus ICV normalised volumes. We propose that a Bayesian approach for longitudinal neuroimaging modelling has merit for providing further under- standing of brain atrophy over time. These views were demonstrated using an application of BLME models applied to a longitudinal AIBL study.

Comparisons of volumetric rate of change of diagnosis level trajectories were compared for HC, MCI and AD participants, with an estimated probability greater than 0.8 on the order of disease pathology. Ranking of converters in retrospect to the study cohort and diagnosis trajectories over age based on volumetric quantiles are the first BLME analysis of their kind applied to longitudinal neuroimaging data. This analysis identified HC to MCI converters most at risk of AD like rate of deterioration and posterior rank distributions provided probabilities on individuals of interest in the worst 15^{th} percentile rank for both regions. The predictive capability of future converters can be derived from these BLME models, as individuals with high neurodegeneration estimates would rank at the extremes in comparison with the remainder of the cohort. The uncertainty of their rank values among a specified quantile is expressed in terms of probabilities, and individuals with a high probability of ranking at extreme levels of neurodegeneration may be indicative of their progressive pathway to further stages of dementia. However to rigorously validate this analysis a richer data set with many more repeated measures and converters over all categories (HC to MCI or AD and MCI to AD) observed at various ages is required. Furthermore, the diagnosis trajectories for each volume region identified critical points in time for both ventricle and hippocampus degeneration from which participants are most likely to show greater deterioration rates. Alternatively a similar analysis can also be used to compare diagnosis trajectories of different longitudinal neuroimgaing population studies, such as the Alzheimer's Disease Neuroimaging Initiative (ADNI).

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Additional analysis regarding group comparisons can be made. For example, similar probabilities for a simulated population mean in comparison to a biologically meaningful constant could also be inferred. An extension to our second analysis to allow population studies to focus on specific participants of interest and monitor their progression rate throughout follow-ups could assist health professionals in making informed choices with regard to patient care. Alternatively HC to AD converters may also be further analysed and ranked with respect to the cohort, to provide further clues as to why these individuals deteriorated so quickly compared with their slower converter counterparts. It is worth- while to note that these inference extensions would not have been possible had we not first attempted the research methods presented in this paper.

A sensitivity analysis with respect to the prior information used in our analysis was conducted on both the ventricle and hippocampus models. This entailed re-running the MCMC sampling technique for each model based on various specifications of the prior information. The subsequent posterior summaries did not vary considerably based on different prior information. Hence, we conjecture that the results are relatively robust to the priors specified in this work.

Despite every precaution taken to provide robust and reliable conclusions from the BLME models, several authors^{7,10} have noted the limitations and disadvantages of Bayesian statistics applied to longitudinal neuroimaging analysis. In particular, drawbacks of Bayesian statistics in the neuroimaging context are discussed at length in Grunkemeier and Payne (2002)⁶³. These include subjective information that can be incorporated in the BLME model specification, in the way in which the prior is specified. Moreover, computational intensity is often far greater in the Bayesian framework than numerical methods employed in a frequentist analysis. In this paper we incorporated vague priors which are semi-conjugate, as we assumed no prior knowledge of the study analysed; the prior specification were a standard choice as suggested in Gelman and Hill (2006)¹⁹. The additional computational insight given. We predict more complex models and future extensions to the methods presented in this paper may result in an increase of computational time, and this will be a factor to consider for future BMLE models.

Extensions to the BLME models presented in this paper include the addition of more covariates to account for trends and variability sources present in gender, genetic factors and additional demographic characteristics which are a few of the key factors known to affect AD onset and disease progression. Furthermore, as the Bayesian framework is ideal for handling complex models such as generalised linear mixed models^{39,64} and spatio-temporal interactions,^{65,66} extensions of this nature will allow for modelling biomarker deterioration rates of multiple brain regions simultaneously over time.

Acknowledgements We wish to thank the Australian Imaging, Biomarkers and Lifestyle Flagship Study of Ageing (<u>http://www.aibl.csiro.au</u>), including all clinicians, scientist, participants and their families.

Contributorship statement: KM and MIC conceived and designed research concept, CCD provided additional suggestions. Statistical analysis and manuscript drafting was performed by MIC. JD and JF were responsible for the acquisition and interpretation of the data. MIC, JF, JMM, CCD, KM and JD participated in critical revision of the manuscript and approved the final manuscript. MIC is responsible for the overall content as the corresponding author.

Competing interests: The authors declare that they have no competing interests.

Funding: This research was jointly funded by an Australian Postgraduate Award (APA), Commonwealth Scientific and Industrial Research Organisation (CSIRO) Digital Productivity and Services Division, and the ARC Centre of Excellence for Mathematical & Statistical Frontiers.

Data sharing statement: Data available from the Australian Imaging, Biomarkers and Lifestyle longitudinal study of ageing (AIBL). This study is funded by the CSIRO and partners. Access to the data is conditional on approval from the AIBL management committee, for guidelines refer to http://aibl.csiro.au/research/support/.

All R code and simulated data for this manuscript is available at https://github.com/MarcelaCespedes/Bayesian_inference_on_neuroimaging.

	BMJ Open	Pa
REFE	ERENCES	
1.	Villemagne, V. L. <i>et al.</i> Amyloid β deposition, neurodegeneration and cognitive decline in sporac Alzheimer's Disease. <i>Lancet Neurology</i> 12 , 357–367 (2013).	dic
2.	Stoessl, A. J. Neuroimaging in the early diagnosis of neurodegenerative disease. <i>Transl.</i> Neurodegener. 1 , (2012).	
3.	Adaszewski, A., Dukart, J., Kherif, F., Frackowiak, R. & Draganski, B. How early can we predict {A}lzheimer's disease using computational anatomy. <i>Neurobiol. Aging</i> 34, 2815–2826 (2013).	
4.	Mattila, J. <i>et al.</i> A disease state fingerprint for evaluation of Alzheimer's disease. <i>J. Alzheimer's L</i> 27, 163–176 (2011).	Dis.
5.	Weiner, M. W. <i>et al.</i> The Alzheimer's Disease Neuroimaging Initiative: a review of papers publish since its inception. <i>Alzheimer's Dement.</i> 9, e111–e194 (2013).	hed
6.	Wang, Y., Resnick, S. M., Davatzikos, C., the Baltimore Longitudinal Study of Aging & the Alzheim Disease Neuroimaging Initiative. Analysis of spatio-temporal brain imaging patterns by Hidden Markov models and serial MRI images. <i>Hum. Brain Mapping</i> 35 , 4777–4794 (2014).	ner's
7.	Bernal-Rusiel, J., Greve, D. N., Reuter, M., Fischl, B. & Sabuncu, M. R. Statistical analysis of longitudinal neuroimage data with Linear Mixed Effects models. <i>Neuroimage</i> 66, 249–260 (2013)	3).
8.	Ziegler, G., Penny, W. D., Ridgway, G. R., Ourselin, S. & Friston, K. J. Estimating anatomical trajectories with Bayesian mixed-effects modeling. <i>Neuroimage</i> 121 , 51–68 (2015).	
9.	Jack, C. R. <i>et al.</i> Evidence for ordering of Alzheimer disease biomarkers. <i>Arch. Neurol.</i> 68, 1526–2 (2011).	1535
10.	Guillaume, B., Hua, X., Thompson, P. M., Waldorp, L. & Nichols, T. E. Fast and accurate modelling longitudinal and repeated measures neuroimaging data. <i>Neuroimage</i> 94, 287–302 (2014).	g of
11.	Holland, D. <i>et al.</i> Rates of decline in Alzheimer disease decrease with age. <i>PLoS One</i> 7 , e42325 (2012).	
12.	Kruschke, J. K. Bayesian estimation supersedes the t test. J. Exp. Psychol. Gen. 142, 573 (2013).	
13.	Pfefferbaum, A. <i>et al.</i> Variation in longitudinal trajectories of regional brain volumes of healthy r and women (ages 10 to 85years) measured with atlas-based parcellation of MRI. <i>Neuroimage</i> 6 176–193 (2013).	men 5,
14.	Jack Jr, C. R. <i>et al.</i> Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. <i>Lancet Neurology</i> 12, 207–216 (2013).	
15.	Oxtoby, N. P. et al. in Bayesian and grAphical Models for Biomedical Imaging 85–94 (Springer, 2	014).

16. Sperling, R. A. *et al.* Toward defining the preclinical stages of Alzheimer's disease: Recommnedations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

BMJ Open

Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.	Enseignement Superieur (ABES) .	3MJ Open: first published as 10.1136/bmjopen-2016-012174 on 7 February 2017. Downloaded from http://bmjopen.bmj.com/ on June 13, 2025 at Agence Bibliographique
--	---------------------------------	---

de l

for Alzheimer's disease. Alzheimer's and Dementia 7, 280–292 (2011).

- 17. Ferreira, L. K. & Busatto, G. F. Neuroimaging in Alzheimer's disease: current role in clinical practice and potential future applications. *Clinics* **66**, 19–24 (2011).
 - 18. Harville, D. A. & Carriquiry, A. L. Classical and Bayesian prediction as applied to an unbalanced mixed linear model. *Biometrics* 987–1003 (1992).
 - 19. Gelman, A. & Hill, J. *Data analysis using regression and multilevel/hierarchical models*. (Cambridge University Press, 2006).
 - 20. Li, R., Stewart, B. & Weiskittel, A. A Bayesian approach for modelling non-linear longitudinal/hierarchical data with random effects in forestry. *Forestry* **85**, 17–25 (2012).
- 21. Friston, K. J. *et al.* Classical and Bayesian Inference in Neuroimaging: Theory. *Neuroimage* **16**, 465–483 (2002).
- 22. Scheltens, P. Advances in Neuroimaging. *Blue Books Neurol.* **30**, 381–410 (2007).
- 23. Risacher, S. L. & Saykin, A. J. Neuroimaging and other biomarkers for Alzheimer's disease: the changing landscape of early detection. *Annu. Rev. Clin. Psychol.* **9**, 621–622 (2013).
- 24. Rodie, M. E., Forbes, K. P. & Muir, K. in *Understanding Differences and Disorders of Sex Development* (*DSD*) (eds. Hiort, O. & Ahmed, S. F.) **27,** 63–75 (Karger Publishers, 2014).
- 25. Sperling, R. A., Karlawish, J. & Johnson, K. A. Preclinical Alzheimer disease the challenges ahead. *Nat. Rev. Neurol.* **9**, 54–58 (2013).
- 26. Ziegler, G., Penny, W. D., Ridgway, G. R., Ourselin, S. & Friston, K. J. Estimating anatomical trajectories with Bayesian mixed-effects modeling. *Neuroimage* **121**, 51–68 (2015).
- 27. Jack, C. R. *et al.* Age-specific population frequencies of cerebral β-amyloidosis and neurodegeneration among people with normal cognitive function aged 50--89 years: a cross-sectional study. *Lancet Neurol.* **13**, 997–1005 (2014).
- 28. Apostolova, L. G. *et al.* Hippocampal atrophy and ventricular enlargement in normal aging, mild cognitive impairment and Alzheimer's disease. *Alzheimer Dis. Assoc. Disord.* **26**, 17 (2012).
- 29. Tang, X., Holland, D., Dale, A. M., Younes, L. & Miller, M. I. Shape abnormalities of subcortical and ventricular structures in mild cognitive impairment and Alzheimer's disease: detecting, quantifying, and predicting. *Hum. Brain Mapp.* (2014).
- 30. Chupin, M. *et al.* Fully automatic hippocampus segmentation and classification in Alzheimer's disease and mild cognitive impairment applied on data from ADNI. *Hippocampus* **19**, 579–587 (2009).
- 31. Jack Jr, C. R. *et al.* Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *Lancet Neurology* **9**, 119–128 (2010).

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

32.	Holland, D. <i>et al.</i> Rates of decline in Alzheimer disease decrease with age. <i>PLoS One</i> 7 , e42325 (2012).
33.	Huijbers, W. <i>et al.</i> Amyloid-β deposition in mild cognitive impairment is associated with increased hippocampal activity, atrophy and clinical progression. <i>Brain</i> 138 , 1023–1035 (2015).
34.	Rowe, C. C. <i>et al.</i> Amyloid imaging results from the Australian Imaging, Biomarker and Lifestyle (AIBL) study of aging. <i>Neurobiol. {A}geing</i> 31, 1275–1283 (2010).
35.	O'Brien, L. M. <i>et al.</i> Statistical adjustments for brain size in volumetric neuroimaging studies: {S}ome practical implications in methods. <i>Psychiatry Res. Neuroimaging</i> 193 , 113–122 (2011).
36.	Voevodskaya, O. <i>et al.</i> The effects of intracranial volume adjustment approaches on multiple regional MRI volumes in healthy aging and {A}lzheimer's disease. <i>Front. Aging Neurosci.</i> 6 , (2014).
37.	Pinheiro, J. C. & Bates, D. M. <i>Mixed-effects models in S and S-PLUS</i> . (Springer, 2000).
38.	McCullagh, P. Generalized linear models. <i>Eur. J. Oper. Res.</i> 16, 285–292 (1984).
39.	Gelman, A. et al. Bayesian data analysis. (CRC press, 2013).
40.	Robert, C. & Casella, G. <i>Monte Carlo Statistical Methods</i> . (Springer New York, 2010). at <https: books.google.com.au="" books?id="PbsCkgAACAAJ"></https:>
41.	Burden, R. & Faires, J. Numerical analysis. (Cengage Learning, 2011).
42.	Adler, J. <i>R in a nutshell: A desktop quick reference</i> . (' O'Reilly Media, Inc.', 2010).
43.	Plummer, M., Best, N., Cowles, K. & Vines, K. CODA: Convergence Diagnosis and Output Analysis for MCMC. <i>R News</i> 6 , 7–11 (2006).
44.	Wickham, H. ggplot2: elegant graphics for data analysis. (Springer New York, 2009). at <http: book="" ggplot2="" had.co.nz=""></http:>
45.	Gelman, A. & Rubin, D. B. Inference from iterative simulation using multiple sequences. <i>Stat. Sci.</i> 7 , 457–472 (1992).
46.	Jack Jr, C. R. <i>et al.</i> Shapes of the trajectories of five major biomarkers of Alzheimer's Disease. <i>Arch.</i> <i>Neurology</i> 7, 856–867 (2012).
47.	Woolrich, M. W., Behrens, T. E. J. & Smith, S. M. Constrained linear basis sets for HRF modelling using Variational Bayes. <i>Neuroimage</i> 21, 1748–1761 (2004).
48.	Bernal-Rusiel, J. L., Reuter, M., Greve, D. N., Fischl, B. & Sabuncu, M. R. Spatiotemporal linear mixed effects modeling for the mass-univariate analysis of longitudinal neuroimage data. <i>Neuroimage</i> 81 , 358–370 (2013).
	For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

2

3 4

5 6

7 8

9

10 11 12

13 14 15

16

17

18 19 20

21

22 23

24 25

26 27

28

29

30 31 32

33

34 35 36

37

38 39 40

41

42 43

44

45 46 47

48

49 50 51

52 53

54 55

56 57

58

59 60

BMJ Open

Ward, A., Arrighi, H. M., Michels, S. & Cedarbaum, J. M. Mild cognitive impairment: disparity of 49. incidence and prevalence estimates. Alzheimer's Dement. 8, 14-21 (2012). 50. Refshauge, A. & Kalisch, D. Dementia in Australia. Australian Institute of Health and Welfare (2012). at <http://www.aihw.gov.au/WorkArea/DownloadAsset.aspx?id=10737422943> 51. Cole, T. J. & Green, P. J. Smoothing reference centile curves: the LMS method and penalized likelihood. Stat. Med. 11, 1305–1319 (1992). 52. Koenker, R. Quantile regression. (Cambridge university press, 2005). 53. Muggeo, V. M. R., Sciandra, M., Tomasello, A. & Calvo, S. Estimating growth charts via nonparametric quantile regression: a practical framework with application in ecology. Environ. Ecol. Stat. 20, 519–531 (2013). 54. Caroli, A. & Frisoni, G. B. The dynamics of Alzheimer's disease biomarkers in the Alzheimer's Disease Neuroimaging cohort. Neurobiol. Aging **31**, 1263–1274 (2010). 55. Leung, K. K. et al. Cerebral atrophy in mild cognitive impairment and Alzheimer's disease. Am. Academy Neurology 80, 648–654 (2013). 56. Antoniano-Villalobos, I., Wade, S. & Walker, S. G. A Bayesian Nonparametric Regression Model with Normalized Weights: A study of Hippocampal Atrophy in Alzheimer's Disease. J. American Statistical Association 109, 477-490 (2014). 57. Caselli, R. J. et al. Longitudinal modeling of age-related memory decline and the APOE-E4 effect. N. Engl. J. Med. 361, 255-263 (2009). 58. Verghese, P. B., Castellano, J. M. & Holtzman, D. M. Apolipoprotein E in Alzheimer's disease and other neurological disorders. Lancet Neurol. 10, 241–252 (2011). 59. Liu, C. C., Kanekiyo, T., Xu, H. & Bu, G. Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy. Nat. Rev. Neurol. 9, 106–118 (2013). 60. Carter, C. L., Resnick, E. M., Mallampalli, M. & Kalbarczyk, A. Sex and gender differences in Alzheimer's disease: recommendations for future research. J. Women's Heal. 21, 1018–1023 (2012). 61. Li, R. & Singh, M. Sex differences in cognitive impairment and Alzheimer's disease. Front. Neuroendocrinol. 35, 385–403 (2014). 62. Hardy, J. A hundred years of Alzheimer's disease research. *Neuron* **52**, 3–13 (2006). Grunkemeier, G. L. & Payne, N. Bayesian analysis: a new statistical paradigm for new technology. 63. Ann. Thorac. Surg. 74, 1901–1908 (2002). 64. Kleinman, K. P. & Ibrahim, J. G. A semi-parametric Bayesian approach to generalized linear mixed models. Stat. Med. 17, 2579–2596 (1998). For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

BMJ Open: first published as 10.1136/bmjopen-2016-012174 on 7 February 2017. Downloaded from http://bmjopen.bmj.com/ on June 13, 2025 at Agence Bibliographique de Enseignement Superieur (ABES)

to text

and

data mining, Al training, and similar technologies

Protected by copyright, including for uses related

BMJ Open

- 65. Alston, C. L., Mengersen, K. L. & Pettitt, A. N. *Case studies in Bayesian statistical modelling and analysis*. (John Wiley & Sons, 2012).
- 66. Banerjee, S., Carlin, B. P. & Gelfand, A. E. *Hierarchical modeling and analysis for spatial data*. (Crc Press, 2014).
- 67. Llano, D. A., Laforet, G. & Devanarayan, V. Derivation of a new ADAS-cog composite using treebased multivariate analysis: prediction of conversion from mild cognitive impairment to Alzheimer disease. *Alzheimer Dis. Assoc. Disord.* **25**, 73–84 (2011).
- 68. Delor, I., Charoin, J. E., Gieschke, R., Retout, S. & Jacqmin, P. Modeling Alzheimer's Disease Progression Using Disease Onset Time and Disease Trajectory Concepts Applied to CDR-SOB Scores From ADNI. *CPT Pharmacometrics Syst. Pharmacol.* **2**, e78 (2013).
- μη. 3-amyloid load a_η. 69. Jack Jr, C. R. *et al.* Brain β-amyloid load apporaches plateau. *Am. Academy Neurology* **80**, 890–896 (2013).



Fig. 1: Percentile ranges of volume across ages 60 to 85 years old, for ventricle (left) and hippocampus (right). Recall region volumes are normalised by the ICV value as they represent a percentage of volume within the intracranial cavity. Ranges up to the 100th percentile henceforth denote the empirical maximum volume for that region. Volume percentiles; 75-100th from blue (0.75) to top dotted line, 50-75th from green (0.25) to blue (0.75) line, 25-50th from red (0.25) to green (0.50) line and 15-25th from black (0.15) to red (0.25) line.

112x55mm (300 x 300 DPI)

BMJ Open: first published as 10.1136/bmjopen-2016-012174 on 7 February 2017. Downloaded from http://bmjopen.bmj.com/ on June 13, 2025 at Agence Bibliographique de l Enseignement Superieur (ABES)

Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

BMJ Open: first published as 10.1136/bmjopen-2016-012174 on 7 February 2017. Downloaded from http://bmjopen.bmj.com/ on June 13, 2025 at Agence Bibliographique de l Enseignement Superieur (ABES)

Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.





Fig. 2: Posterior densities of population mean estimates of linear deterioration rate for diagnosis (top plot): HC, MCI and AD, for ventricle (left) and hippocampus volume (right) models. Dotted lines on bottom plots denote the means for each density, whose values are shown in Table 1.

190x173mm (300 x 300 DPI)





Fig. 4: Box plots of posterior distribution of random effect values for participants in AIBL study (N = 260) for full data (four time points). Ventricle (top) and hippocampus (bottom) rates of deterioration for each participant in the study. As there are 157 HC, 34 MCI, 42 AD and 27 Converters in this study, there is higher uncertainty on the rate of deterioration of converters, MCI and AD participants (hence longer box plots) as compared to the HC (narrower box plots). Eight individuals who converted from HC to MCI throughout the study are highlighted in red with corresponding ID numbers.

272x246mm (300 x 300 DPI)



Fig. 5: Probability curves show the posterior probability of HC, MCI or AD diagnosis for the ventricle (top) and hippocampus (bottom) models, 95% interval denotes Monte Carlo error based on several simulations of the BLME models. Total volume divided into four percentile volume ranges, as shown in Fig. 1. Percentiles; 75- 100^th,50-75th,25- 50th and 15- 25th.

155x345mm (300 x 300 DPI)

BMJ Open: first published as 10.1136/bmjopen-2016-012174 on 7 February 2017. Downloaded from http://bmjopen.bmj.com/ on June 13, 2025 at Agence Bibliographique de

Enseignement Superieur (ABES)

Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies

Supplementary Material

Please refer to website

https://github.com/MarcelaCespedes/Bayesian_inference_on_neuroimaging for full R code (including code for plots) used in analysis outlined in the manuscript *Comparisons of neurodegeneration over time between healthy ageing and Alzheimer's disease cohorts via Bayesian inference.*

1 Posterior Predictive checks and parameter estimates

Posterior predictive checks were carried out to assess goodness of fit and prediction capability of our models in expression (3) of the manuscript, as predicted values were simulated from the joint posterior distribution. After burn-in and thinning, as specified in Section 3.2 of the manuscript, each predicted value consists of 8,000 simulations from which we compute the 95% credible intervals. Posterior predictive plots are shown in Figure S1. MCMC chain diagnostics such as trace, density and auto-correlation plots as well as the Gelman and Rubin convergence measures are available upon request.



Figure S1: Posterior predictive means versus response values with the 95% credible interval. The tight bandwidth on all responses shows we have adequately captured the variability. As both the plots show a general diagonal pattern of x = y for majority of the values (with the exception of a few cases), this provides evidence of accurate predicted values from our model.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
20
22
22
23
24
20
20
21
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
50

	Ventricle ESS	Hippocampus ESS
β_{0i}	7050	6709
β_{1i}	7101	3388
β_2	5152	3164
β_3	3817	1932
β_{4i}	5140	1390
β_{5i}	5252	946
σ^2	8000	2963
σ_0^2	8000	2798
σ_1^2	7562	1820
$\sigma_4^{\overline{2}}$	6614	845
σ_5^{2}	5599	1278
DIC	- 5822	-4271
P.P.	0.993	0.995

Table S1: Posterior proportion of response (P.P), is a proportion of predicted values which lie within 95% credible interval of prediction values as seen in Figure S1. Effective sample size (ESS) denotes the estimated number of independent samples (no auto-correlation) obtained in our estimated parameters. As per our burn-in and thinning specifications stated in Section 3.2 of the manuscript, the ESS will be at most a value up to 8,000. Deviance information criterion (DIC) is an overall measure of goodness of fit as described in Section 3.2.

Distribution of ranks for converters 2

As described in Section 4.2 of the manuscript, distribution of ranks were performed on all (27) converters of the AIBL study, first on a subset of the first three time points; for ventricle model see Figure S2, hippocampus see Figure S3. Similarly the distribution ranks were estimated on the whole data set, Figure S4 shows the results for the ventricle model, and Figure S5 correspond to the hippocampus model.



Figure S2: Ventricle converters posterior distribution of ranks for the first three time points.





Figure S3: Hippocampus converters posterior distribution of ranks for the first three time points.



Figure S4: Ventricle converters posterior distribution of ranks for full data (4 timepoints).



Figure S5: Hippocampus converters posterior distribution of ranks for full data (4 timepoints).

3 APOE and Gender diagnosis trajectories over age

As mentioned in Section 4.3 of the manuscript, initial exploration of diagnosis trajectories over groups; male, female, apolipoprotein $\varepsilon 4$ (APOE $\varepsilon 4$) carriers and non-carriers were also investigated for the ventricle and hippocampus models.

The broad prevalence rates utelised for Inference 3 were derived from Ward et al. (2012); Refshauge and Kalisch (2012) and is summarised in Table S2. Again the reader is cautioned that these are very broad estimates of prevalence rates and are generalised over many factors including lifestyle, genetic and demographic. These prevalence rates also do not take into account participants who develop other forms of dementia or any other neuropsychological disorders. The authors acknowledge there are several factors which the models presented in the manuscript do not account for. As the BLME models and inference derivation presented in this paper are the first of its kind, the objective of Inference 3 is to demonstrate probable diagnosis trajectories conditional on very broad, non-group specific prevalence rates. In order to account for gender and APOE $\varepsilon 4$ status and develop diagnosis trajectories specific to these groups, prevalence rates across ages 65-85 specific to these groups is required, which unfortunately is difficult to find in literature. Figure S6 are the disease trajectories for models (3) in the manuscript applied on male, female, APOE $\varepsilon 4$ carriers and non carriers groups separately, for the ventricle and hippocampus models. We assumed the same prevalence rates as in the manuscript.

Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.	MI Open: first nublished on 10.96/kmisnen 2016 012121 on 7 Eckney 2017 Develoaded from http://kmisnen.kmis.com/on lung 12, 2025 of Acapas Bibliographic
--	---

Age	HC	MCI	AD
60	0.945	0.037	0.018
65	0.917	0.055	0.028
70	0.859	0.096	0.045
75	0.592	0.333	0.075
80	0.518	0.357	0.125
85	0.466	0.301	0.203

Table S2: Broad prevalence rates for healthy control (HC), mild cognitive impaired (MCI) and Alzheimer's disease taken from Ward et al. (2012); Refshauge and Kalisch (2012). These rates do not account for any lifestyle, demographic and genetic factors as well as other forms of dementia and neuropsychological disorders which are known to affect prevalence rates.

References

A.	Refshauge	and	D.	Kalisch.	Dementia	in	Australia,	2012.	URL
ł	<pre>http://www.</pre>	aihw.	gov.au	J/WorkArea ا	/DownloadAsse	et.as	px?id=107	37422943.	Cat.
n	o. AGE70.								

A. Ward, H. M. Arrighi, S. Michels, and J. M. Cedarbaum. Mild cognitive impairment: disparity of incidence and prevalence estimates. *Alzheimer's & Dementia*, 8(1):14–21, 2012.





Figure S6: Male, female, APOE ε 4 carriers and non-carriers diagnosis trajectories for ventricle (top) and hippocampus (bottom) model. Volume quantiles X1, X2, X3 and X4 denote 75-100th, 50-75th, 25-50th and 15-25th quantiles respectively. 6

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

HC HC MCI AD

BMJ Open

Comparisons of neurodegeneration over time between healthy ageing and Alzheimer's disease cohorts via Bayesian inference

Journal:	BMJ Open
Manuscript ID	bmjopen-2016-012174.R1
Article Type:	Research
Date Submitted by the Author:	26-Oct-2016
Complete List of Authors:	Cespedes, Marcela; Queensland University of Technology, School of Mathematical Sciences Fripp, Jurgen; CSIRO Digital Productivity and Services, Australian E-Health Research Centre McGree, James; Queensland University of Technology, School of Mathematical Sciences Drovandi, Christopher; Queensland University of Technology, School of Mathematical Sciences Mengersen, Kerrie; Queensland University of Technology, School of Mathematical Sciences Doecke, James; CSIRO Digital Productivity and Services, Australian E- Health Research Centre
Primary Subject Heading :	Neurology
Secondary Subject Heading:	Neurology, Epidemiology, Radiology and imaging
Keywords:	Bayesian inference, mixed effects models, neurodegeneration, Alzheimer's disease, longitudinal neuroimaging study

SCHOLARONE[™] Manuscripts 0

Comparisons of neurodegeneration over time between healthy ageing and Alzheimer's disease cohorts via Bayesian inference

Corresponding author

Marcela Ines Cespedes

School of Mathematical Sciences, Queensland University of Technology GPO BOX 2434 Brisbane QLD 4001 Australia Marcela.Cespedes@hdr.qut.edu.au Tel: +61-07-313-87440

Co-authors:

3

- 1 Jurgen Fripp CSIRO Digital Productivity and Services Australia E-Health Research Centre Herston, QLD, Australia
- 2 James M. McGree School of Mathematical Sciences Queensland University of Technology Brisbane, Australia
 - Christopher C. Drovandi School of Mathematical Sciences Queensland University of Technology Brisbane, Australia
- 4 Kerrie Mengersen School of Mathematical Sciences Queensland University of Technology Brisbane, Australia
- 5 James Doecke CSIRO Digital Productivity and Services Australia E-Health Research Centre Herston, QLD, Australia

Keywords: Bayesian inference; mixed effects models; neurodegeneration; longitudinal neuroimaging study; Alzheimer's disease

Manuscript word count: 6,500 words

BMJ Open: first published as 10.1136/bmjopen-2016-012174 on 7 February 2017. Downloaded from Protected by copyright, including for uses related Enseignement Super ő tp://bmjopen.bmj.com/ on June 13, 2025 at Agence Bibliographique de Al training, and similar technologies

Comparisons of neurodegeneration over time between healthy ageing and Alzheimer's disease cohorts via Bayesian inference

Marcela I. Cespedes · Jurgen Fripp · James M. McGree · Christopher C. Drovandi · Kerrie Mengersen & James D. Doecke

Abstract

Objectives: In recent years, large scale longitudinal neuroimaging studies have improved our understanding of healthy ageing and pathologies including Alzheimer's disease (AD). A particular focus of these studies is group differences and identification of subjects at risk of deteriorating to a worse diagnosis. For this, statistical analysis using Linear mixed effects (LME) models are used to account for correlated observations from individuals measured over time. A Bayesian framework for LME models in AD is introduced in this paper to provide additional insight often not found in current LME volumetric analyses.

Setting and participants: Longitudinal neuroimaging case study of ageing were analysed in this research on 260 participants diagnosed as either healthy controls (HC), mild cognitive impaired (MCI) or AD. Bayesian LME models for the ventricle and hippocampus regions were used to; (i) estimate how the volumes of these regions change over time by diagnosis, (ii) identify high risk non-AD individuals with AD like degeneration, and (iii) determine probabilistic trajectories of diagnosis groups over age.

Results: We observed (i) large differences in average rate of change of volume for the ventricle and hippocampus regions between diagnosis groups, (ii) high risk individuals who had progressed from HC to MCI and displayed similar rates of deterioration as AD counterparts, and (iii) critical time points which indicate where deterioration of regions begin to diverge between the diagnosis groups.

Conclusion: To our knowledge, this is the first application of Bayesian LME models to neuroimaging data which provides inference on a population and individual level in the AD field. The application of a Bayesian LME framework allows for additional information to be extracted from longitudinal studies. This provides health professionals with valuable information of neurodegeneration stages, and a potential to provide a better understanding of disease pathology.

Strengths of this study

- The models presented in this research address realistic challenges in a longitudinal study setting such as; large patient drop-out (unbalanced design), large and small diagnosis groups and noisy MRI observations.
- This is the first study of its kind to incorporate data external to this analysis, in terms of prevalence rates, in conjunction with the statistical models to infer disease trajectories for brain regions over age.

Limitations of this study

- This research does not accommodate for participants with other neurological disorders and assumes participants are in one of three groups; healthy control (HC), mild cognitive impaired (MCI) and Alzheimer's disease (AD).
- es which ... atic status. • Additional covariates which are known to affect neurodegeneration were not included in this analysis, such as gender and genetic status.

1 INTRODUCTION

Alzheimer's disease (AD) is the most common form of dementia worldwide¹. Advances of neuroimaging techniques have been useful for early diagnosis of neurodegenerative disorders ²³ and, coupled with mathematical and statistical models, provide insight to better understand healthy ageing and disease pathology degeneration.^{4–6} The use of linear mixed effects (LME) models has been advocated by Bernal- Rusiel et al. (2013a)⁷ and more recently by Ziegler et al. (2015)⁸ to characterise longitudinal degeneration from neuroimaging data. Bayesian LME (BLME) models are applied in this research to provide insight into the diagnosis of AD over time. In this research we address three main areas; population diagnosis comparisons based on estimated volumetric rate of change over age, ranking of participants by order of linear volumetric rate of change, and region specific probability trajectories across age of diagnosis groups, conditional on prevalence rates.

Recent state-of-the-art analysis on clinical diagnosis classification groups emphasise the need to better understand disease pathology in asymptomatic and early stages of AD individuals.^{9–13} A strong focus of longitudinal neuroimaging studies is to monitor morphological changes among healthy control (HC), mild cognitive impaired (MCI) and AD groups as they progress throughout the disease continuum.^{7,14}

Previous LME models of volumetric degeneration reported on comparisons assessing ranking of diagnosis levels.^{7,15} However, in these studies, the magnitude of the differences of disease progression as well as their estimated variances is often excluded,^{7,14–17} thus a richer insight into the differences of diagnosis levels is lacking. The BLME approach uses simulation techniques to draw from the posterior distribution, which is a combination of prior information and information from the data (through the likelihood function), to provide diagnosis group estimates and comparisons. These simulations quantify uncertainty and provide posterior probabilities that can be compared directly, without referring to significance levels or multiple statistical tests.

The development of methods which account for large inter and intra variability of biomarkers presents a challenge in longitudinal neuroimaging studies.^{18–20} Furthermore, the observations of diagnosis group tends to become unbalanced over time which makes it difficult to deduce information of the complex AD pathway. However, insight into neurodegeneration of high risk participants, namely MCI, is crucial for early detection methods and improving diagnostic accuracy of AD.^{5,21} Several authors such as Harville and Carriquiry (1992),²² Gelman and Hill (2006)²³ and Li et al. (2012)²⁴ state BLME models have the capability to seamlessly handle unbalanced data and small-sample design analysis. This motivates our choice of statistical framework, as we aim to utilise as much information as possible from the study analysed, and retain participants with a single observation.

Individuals ranked by order of neurodegeneration severity allows for comparisons of progression of all individuals over the study, while quantifying the uncertainty and estimating variability of individualised conversion rates. The application of BLME models allows for estimation of class membership probabilities and estimation of deterioration rates of each participant via the analysis of random effects. This type of analysis is often overlooked in longitudinal studies of ageing.²⁵

As the field of neuroimaging in AD has been rapidly expanding in the past 20 years,^{26–29} it is of interest to incorporate as much relevant information as possible, as independent longitudinal neuroimaging studies often build on and support each other.^{5,30,31} This can be achieved using the Bayesian approach, as it combines external information with experimental data at hand, while accounting for various sources of uncertainty. Often this background information can be incorporated in the form of the prior, but it can also be applied after estimation of the model to provide additional inference from our model outcomes. In the current project, we demonstrate this concept by combining model information with prior knowledge obtained from prevalence studies to formulate probabilistic diagnosis group trajectories over age.

Authors Jack et al. $(2014)^{32}$ highlight the importance of population frequency or probabilistic trajectories of neurodegeneration groups over a wide age span. Their study quantified frequencies of expected neurodegeneration cases dependent on ages 50 to 89. Particular focus was placed on asymptomatic individuals (preclinical AD) who were at risk of developing AD and ages of increased frequency of convergence to AD as they reach their later years. While our methods can also be used for similar purposes and place emphasis on a particular neurodegeneration group, the goal for our final analysis is to identify critical time points where all diagnosis levels begin to diverge. This can aid in discovering groups or patterns in neurodegeneration consistent with healthy ageing

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

BMJ Open

or the AD pathway. Alternatively a similar analysis can also be used to compare diagnosis trajectories of different longitudinal neuroimgaing population studies, such as the Alzheimer's Disease Neuroimaging Initiative (ADNI).

This paper is outlined as follows. Section 2 describes the case study, Sections 3.3 and 4.1 show an application of the BLME models to address multiple comparisons of various sizes from baseline diagnosis, including large $(N_{HC} = 168 \text{ people})$ and small groups $(N_{MCI} = 50 \text{ and } N_{AD} = 42 \text{ people at baseline})$. Sections 3.4 and 4.2 rank individuals by order of neurodegeneration severity, thereby comparing the progression of all individuals over the study time. Approximately 10% of individuals convert from baseline case to a worse diagnosis throughout the length of the study. This analysis allows for the identification of those participants who are most at risk of developing AD like rates of deterioration for the hippocampus and ventricle regions of the brain. The third and final area addressed in this research is presented in Sections 3.5 and 4.3, which estimates probabilistic diagnosis group trajectories across age, derived from neuroimaging information. This requires the synthesis of information for the study cohort and the AD literature.

2 AIBL LONGITUDINAL STUDY OF AGEING

The neuroimaging data analysed in this paper were obtained from the Australian Imaging Biomarker and Lifestyle Study of Ageing (AIBL). This is an ongoing study which aims to discover which biomarkers such as cognitive assessment results, neuroimaging, lifestyle and demographic factors potentially influence subsequent development of AD. The sample comprises N = 260 people, who have at most four repeated observations approximately 18 months apart. These data are highly unbalanced, since patient drop out occurs at every time point throughout the study, with approximately 69% of participants in the final follow up.

Key regions of the brain which are strongly associated with neurodegeneration in relation to AD and healthy ageing include the lateral ventricles^{33,34} and hippocampus volumes.^{3,15,35–37} Atrophy due to disease pathology spreads throughout particular regions such as the hippocampus, which leads to a general decrease in volume over time. The decrease in brain matter results in an increase of cerebrospinal fluid (CSF) which bathes and cushions the brain and spinal cord. The lateral ventricles are filled with CSF, hence an increase of overall brain atrophy results in an increase of ventricle volume. Models presented here were considered separately for the lateral ventricles and the sum of the left and right hippocampal (hippocampus) volume derived from MRI. See Rowe et al. (2010)³⁸ for details on image acquisition and processing. While we cannot deduce entire brain neurodegeneration inferences from the analysis of two regions, in this research we discuss in detail the application of two well-known AD related regions, and note, the BLME models presented here can be easily applied to any other region of interest.

Brain region volumes were normalised by the intracranial volume (ICV), hence all volumes are in the (0, 1) interval. This accounts for the variability of different cranial sizes, while preserving the trend in volume.^{39,40} Due to the wide range in values and in order to eliminate numerical problems in the estimation of these models, age was standardised $(age - \overline{age})/sd(age)$, where \overline{age} and sd(age) are the empirical mean and standard deviations of the study group ages. Likewise, the hippocampus ICV response was scaled up by a factor of 100, in order to avoid variance estimates close to zero which can be difficult to estimate. All participants in this study were categorised as: healthy control (HC), mild cognitive impaired (MCI) and those with a probable diagnosis of Alzheimer's disease (AD) at each time point based on neuropsychological diagnosis. The aim of the BLME's was to capture the linear decrease in regional brain volume across ages for people within three diagnosis groups.

3 METHODS

LME models are a standard approach to modelling repeated observations from several individuals.⁴¹ Standard LME models require the following assumptions to be met: a linear relationship exists between the response and the explanatory variables; the terms at every level are Gaussian although for non-normal models we may extend this assumption to the exponential family and apply generalised linear mixed models;⁴² the variances across all levels are homoscedastic, and repeated observations for an individual can be correlated, but observations between people are assumed independent. The general LME model is of the following form,

 $y = X\beta + Zb + \varepsilon$

where X and Z denote design matrices, and vectors $\boldsymbol{\beta}$ and \boldsymbol{b} are the fixed and random effects respectively for r fixed, m random effects and a total sample size of n observations. The residual vector $\boldsymbol{\varepsilon}$ is assumed to be normally distributed with $\boldsymbol{\varepsilon} \sim MVN(0, \sigma^2 I_n)$, where I_n is the $n \times n$ identity matrix. While our response values are constrained to

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

(1)

BMJ Open: first published as 10.1136/bmjopen-2016-012174 on 7 February 2017. Downloaded from http://bmjopen.bmj.com/ on June 13, 2025 at Agence Bibliographique de Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

the (0, 1) range the assumptions of the model were assessed via a histogram of the residuals, scatter and quantile-quantile plots and were found to not deviate from our model assumptions, refer to the Supplementary material. The parameters in this analysis are in the volume ICV/ Standard age unit and careful back transformation is required to convert to an alternative unit, such as mm³/ year. The random effects vector \boldsymbol{b} is assumed to be multivariate normally distributed, $\boldsymbol{b} \sim MVN(\boldsymbol{0}, \Sigma)$, where the variance covariance matrix of the random effects is denoted by Σ .

3.1 Statistical analysis

In a Bayesian framework the likelihood corresponding to the model in equation (1) is $p(\mathbf{y}|\boldsymbol{\beta}, X, Z, \boldsymbol{b}, \sigma^2, \Sigma)$, which is conditional on the random effects and on the model parameters. The resultant joint posterior distribution for the model parameters and random effects given the data is

$$p(\boldsymbol{\beta}, \boldsymbol{b}, \sigma^2, \boldsymbol{\Sigma} | \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{y}) \propto p(\boldsymbol{y} | \boldsymbol{\beta}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{b}, \sigma^2, \boldsymbol{\Sigma}) p(\boldsymbol{b}, \boldsymbol{\Sigma}) p(\boldsymbol{\beta}) p(\sigma^2) p(\boldsymbol{\Sigma}).$$
(2)

In the absence of external information, weakly informative priors, $p(\boldsymbol{\beta}), p(\sigma^2)$ and $p(\Sigma)$ were used throughout; refer to expression (3) in Section 3.2 for full specification of priors. Under the Bayesian paradigm all the assumptions stated in Section 3 remain. Furthermore as Gelman et al. (2013)⁴³ and Gelman and Hill (2006)²³ state, additional complexity and generalisation of the LME model comes naturally under the Bayesian framework.

Estimation of the model parameters was achieved by sampling from the joint posterior distribution using Markov chain Monte Carlo (MCMC) techniques⁴⁴ which samples from the marginal posterior distributions as a by-product. Note that the parameter estimates are obtained by integrating over the posterior distribution, rather than maximising the likelihood, as numerical methods to solve integrals in high dimensions are often difficult to compute.^{44,45}

3.2 BLME in the context of the case study

Following equation (1), the normalised volume is denoted by Y_{ij} for the i^{th} individual at the j^{th} time point, where binary values x_{MCI} and x_{AD} refer to the two levels of diagnosis, MCI and AD respectively with HC as the baseline. The Bayesian LME model for person i = 1, 2, ..., 260 at time point j = 1, 2, 3, 4 is given by

 $Y_{ij} \mid \mu_{ij}, \sigma^{2} \sim N(\mu_{ij}, \sigma^{2})$ $\mu_{ij} = \beta_{0} + \beta_{1} x_{MCI,ij} + \beta_{2} x_{AD,ij} + \beta_{3i} StndAge_{ij} + \beta_{4i} StndAge_{ij} x_{MCI,ij} + \beta_{5i} StndAge_{ij} x_{AD,ij}$ $\beta_{ki} = \beta_{k} + b_{ki}, \quad for \ k = 0, 3, 4, 5$ $\boldsymbol{b}_{i} \sim MVN(\boldsymbol{0}, \Sigma).$ (3)

Random effects $\mathbf{b}_i = [b_{0i}, b_{3i}, b_{4i}, b_{5i}]$ denotes the i^{th} individuals deviation from population means $\beta_0, \beta_3, \beta_4$ and β_5 . The model in (3) allows for correlation between random effects and this is reflected by the structure of the priors. The residual variance and the variance-covariance matrices are designated by semi-conjugate priors $\sigma^2 \sim IG(0.001, 0.001)$ and $\Sigma \sim Wishart(R, 6)$ respectively, where $= 1,000 \times I_4$. The fixed effects vector $\boldsymbol{\beta} = [\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5]^T$ is assumed normal with $\boldsymbol{\beta} \sim MVN(\mathbf{0}, 1e6, \times I_6)$. Non-linear trends in Age were investigated in order to derive an appropriate model for our application, (refer to Supplementary material for further details). However, the linear predictor in expression (3) was found to approximately represent the data. Posterior predictive plots were used as a measure of goodness-of-fit. This involved simulating from the posterior distribution and forming 95% credible intervals of the posterior predictive responses, which were compared with the observed responses.

The R software was used to implement the Bayesian models.⁴⁶ The rjags package⁴⁷ implemented MCMC methods to estimate the parameters. Packages $coda^{47}$ and $ggplot2^{48}$ were used to analyse the MCMC chains and visualise the three sets of analyses presented here. All R source code for this manuscript and simulated data is available at github website <u>https://github.com/MarcelaCespedes/Bayesian inference on neuroimaging</u>.

Two independent MCMC runs were performed using different starting values; each chain ran for 300K iterations of which 100K were discarded as burn-in and the remaining simulations were thinned at every 50th iteration. The retained 8,000 simulations were taken as samples from the posterior distribution. Convergence diagnostics of the chains included observing the trace, density and autocorrelation plots as well as the Gelman and Rubin diagnostic.⁴⁹ Desirable chain mixing and convergence was observed in all diagnostics. In addition to the residual For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

and posterior checks, leave-one-out cross validation (LOOCV) was performed to assess the models predictive capability of new data, and the mean squared error (MSE) was computed on both models. In a hierarchical setting, the size of the data and how balanced the structure is heavily affects the relative performance of the model.⁵⁰ For this reason we performed two approaches for LOOCV on the ventricle and hippocampus models. First all the observations for an individual were omitted from the analysis (and therefore all of their data), and this was repeated for all individuals. Secondly, for those participants with more than one observation (199 participants in our data set), a single observation was randomly removed from the analysis, refer to the Supplementary material for full results. In practice we wish to minimise the MSE, as it comprises of the sum of the variance, bias squared and irreducible error. Both LOOCV approaches demonstrated low MSE values, which supports our model choice, refer to Supplementary material for full details.

For comparison, the research questions addressed here were attempted with model (3) fitted in the classical framework for both regions. Sections 3.5, 4.2, 4.3 discuss the results for each analysis

3.3 How do HC, MCI and AD participants degenerate over time?

Performing a Bayesian analysis provides a posterior distribution of the parameter which here can be used to estimate the rate of volumetric degeneration for each diagnosis level.¹⁶ In this analysis, we estimate a diagnosis group effect via the posterior mean of the relevant parameter, and investigate differences in these effects via credible intervals (about differences of these means). Other than mean diagnosis comparisons, further analysis in terms of mean differences of these groups is often not performed in LME volumetric neuroimaging models.^{7,51} However as highlighted in Apostolova et al. (2012)³³ and Holland et al. (2012)¹⁵, such insight allows for potential techniques to detect signs of AD like neurodegeneration on pre-symptomatic individuals.

As indicated in (3), the population rate of deterioration for each diagnosis consists of the addition of the baseline effect (HC) with the interaction terms for the other diagnosis groups (MCI or AD). Thus the posterior marginal distributions of β_3 for the baseline, $\beta_3 + \beta_4$, $\beta_3 + \beta_5$ for MCI and AD diagnosis respectively, were compared.

Furthermore, the order of deterioration of the diagnosis levels over both brain regions was assessed. Posterior probabilities were used to order parameter values, since this allows for direct probabilistic diagnosis group comparisons based on the MCMC output while quantifying uncertainty in the parameter estimates. Let M be the number of MCMC posterior draws; in our methods M = 8,000 as described in Section 3.2.

The probability that the rate of change for MCI degeneration is smaller than an AD diagnosis for the ventricle region is estimated by

$$P(MCI < AD) = \frac{1}{M} \sum_{m=1}^{M} \mathbb{1}(\beta_4^m - \beta_5^m < 0),$$
⁽⁴⁾

where the indicator function 1 is equal to 1 if $\beta_4^m - \beta_5^m < 0$ and 0 otherwise. Probabilities for other comparisons of diagnosis levels for the ventricle and hippocampus regions are computed in a similar manner; see Section 4.1 for full results.

3.4 How to identify individuals with high levels of neurodegeneration?

It is expected that individuals who are healthy (HC) will have relatively minimal deterioration while those with MCI or AD will show increasing levels of deterioration. Hence we would expect that the volumetric rate of change will reflect the neuropsychological clinical diagnosis. However as noted by Woolrich et al. (2004);⁵² Bernal-Rusiel et al. (2013a)⁷ and Bernal-Rusiel et al. (2013b)⁵³, high inter and intra variability is often observed in longitudinal neuroimaging studies. For this reason, in this analysis we foresee the estimated volumetric rate of change for a few individuals not to group with participants of the same diagnosis and exercise caution when comparing estimated trajectories of individuals with a single observation.

Participants with outlier rates of deterioration or not within range of their diagnosis levels, as well as those who converted throughout the study are of particular interest as they do not conform to the overall trend over time ordering. Thus a question of interest might be: if an individual has a high neurodegeneration rate with respect to their corresponding diagnosis group, are they likely to degenerate along the AD pathway?

In our data, one individual progressed directly from HC to AD, two were observed to follow the full spectrum (HC to MCI to AD throughout all four follow-ups), eight people progressed from HC to MCI and a further 16 individuals For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

progressed from MCI to AD. These converters can be tracked to observe their severity with respect to the rest of the cohort. In this section, particular focus is on the converters who progressed from HC to MCI, and the comparison of their estimated rates of deterioration with AD participants, as they could be potential AD converters and estimate their probability of remaining in such high rank.

Unlike our first analysis, which compared the estimated population effect across all diagnosis levels, the focus here is on an individual's rate of deterioration. The marginal posterior distributions of individual random effects values in the HC (β_{3i}), MCI ($\beta_{3i} + \beta_{4i}$) and AD ($\beta_{3i} + \beta_{5i}$) groups are inspected, to estimate the rate of deterioration, for i = 1, 2, ..., 260 individuals on all four time points.

Furthermore, as discussed in Section 3.4, the ordered box plots in Fig. 4 illustrate median rankings of participants and illustrate the large variation between individuals. Distribution of ranks on participants take into account the high variation between individuals, by ranking participants at every iteration of the MCMC simulation of the random effects. This results in M = 8,000 simulations on every individual and allow us to derive probabilistic statements on individuals of interest remaining in a specified ranking range, for example the top 15th quantile. This analysis was performed on both a subset of the data, using observations with the first three time points as well as on the full data (four time points) to investigate the change of rank probabilities over time for particular individuals of interest. Such analysis extends the BLME models to allow the identification of high risk converters among the participants analysed. Full results are described in Section 4.2.

3.5 How do diagnosis trajectories vary over age?

The ventricle and hippocampus models derived in (3) were used to compute probabilities $P(\tilde{y}|HC, age)$, $P(\tilde{\gamma}|MCI, age)$ and $P(\tilde{\gamma}|AD, age)$, for a specified age with volume range denoted by $\tilde{\gamma}$. Given information available on an individual at an early age and within the limits of our data age span, we seek to answer: at this early age, for a given volume range, what is the probability that this new individual will be diagnosed as HC, MCI or AD? Moreover, how does this change as the individual ages? These probabilities are estimated below.

At a given age for both ventricle and hippocampus models stated in (3) with diagnosis levels Diagnosis ={*HC*, *MCI*, *AD*}, the following holds

$$P(AD|\tilde{y}, age) = \frac{P(\tilde{y}|AD, age)P(AD|age)}{\sum_{d=1}^{3} P(\tilde{y}|Diagnosis_d, age)P(Diagnosis_d|age)}$$

The BLME model estimates $P(\tilde{y}|AD, age)$, $P(\tilde{y}|MCI, age)$ and $P(\tilde{y}|HC, age)$. As M is the number of MCMC posterior draws, then

$$P(\tilde{y}|AD, age) = \frac{1}{M} \sum_{m=1}^{M} \mathbb{1}(\tilde{y}_m \in \tilde{y}),$$

where the indicator function 1 is equal to 1 if $\tilde{y}_m \in \tilde{y}$ and 0 otherwise. This expression is the average number of predicted values \tilde{y}_m which fall within \tilde{y} . A similar expression was used for MCI and HC diagnosis levels. Probabilities P(HC|age), P(MCI|age) and P(AD|age) were obtained from Ward et al. (2012)⁵⁴ and Refshauge and Kalisch (2012)⁵⁵ for ages 60, 65, 70, 75, 80 and 85. We acknowledged that these are very broad estimates which are generalised over genders, genetic status and many other factors which are known to affect prevalence rates. These prevalence rates also do not take into account participants who develop other forms of dementia, or any other neuropsychological disorders. Refer to the Supplementary material for the full table of probabilities used in this analysis. Similar computations were performed for the other diagnosis levels, MCI and HC, to evaluate related probabilities. Due to the wide variability observed in the hippocampus and ventricle volumes among participants, the volume regions were divided into four different ranges, \tilde{y} , which vary over age groups. Quantile growth curves discussed in Cole and Green (1992)⁵⁶, and Koenker (2005)⁵⁷, highlight the advantages of algorithms that can estimate non-crossing quantiles which are monotone increasing over age to reflect the heteroscedasticity often found in biological systems. In this paper we utilised the algorithm discussed in Muggeo et al. (2013)⁵⁸, as it addresses all of these issues and is available via R package quantregGrowth. The \tilde{y} values of took on ranges; $75 - 100^{th}$, $50 - 100^$ 75^{th} , $25 - 50^{th}$ and $15 - 25^{th}$ percentile of observed response values, as shown in Fig. 1.

*** Figure 1 ***

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

7

1

2 3

4

5

6

7

8

9 10

11

12

13

14

58 59

60

(6)

(5)

For completeness in our analysis, volume ranges such as $5 - 25^{th}$ percentile were explored. However there was very little difference in the probability trajectories among these volume ranges, hence we maintained the $15 - 25^{th}$ percentile range. Furthermore, we wished to avoid low volume outliers, and place emphasis on the degenerating trends present in the majority of the data, for biologically meaningful inferences.

The results from applying Eq. (5) show probability trajectories of an individual being in one of the three diagnosis levels, across ages 60-85 within the four quantile ranges. The goal for this analysis is to identify critical time points where diagnosis levels begin to diverge which can aid in discovering groups or patterns in neurodegeneration consistent with healthy ageing or the AD pathway. Furthermore, the influence of covariates gender and apolipoprotein-E (APOE) was explored, by repeating this analysis on sub-groups of male, female, APOE positive and negative.

A similar analysis cannot be performed with a classical LME model, as the method of maximation of the likelihood does not allow for the straightforward computation of probabilities $P(HC|\tilde{y}, age), P(MCI|\tilde{y}, age)$ and $P(AD|\tilde{y}, age)$. Another drawback of the classical approach is that it does not lend itself to the incorporation of relevant additional external data, to further extend statistical inference.

4 RESULTS

4.1 How do HC, MCI and AD participants degenerate over time?

The atrophy patterns for the ventricle and hippocampus regions described in Section 2 are reflected in the results of the BLME models. A decrease of hippocampus volume and an increase of ventricle volume is depicted by the posterior densities for the rates of deterioration for the two responses as shown in Fig. 2. As expected, this biological pattern across the three levels of diagnosis is reflected in Fig. 2 as well as in Tables 1 and 2. The ventricle population estimates of deterioration show an increase of volume as the diagnosis progressively worsens and the hippocampus population estimates of deterioration reflect a decreasing negative slope from HC, to MCI and AD. The overlapping densities are expected as individuals generally progress gradually in order of deterioration, from HC to MCI to AD. Despite this overlap, there are distinct differences between the average rate of volumetric deterioration between the three diagnoses, as seen in Table 1.

*** Figure 2 ***

Tables 1 and 2 present estimated rates of change as well as the probabilities of diagnosis ordering for the hippocampus and ventricles. Furthermore, the difference among HC and degeneration levels MCI and AD, show the additional annual standardised age rate of change from baseline. The increasing range of the credible intervals for each group as deterioration progresses from HC to MCI to AD, illustrate the stratified structure of different sample sizes over groups in our data. The box plots in Fig. 4 also demonstrates the general variability due to various diagnosis sample numbers

Table 1: Top: Posterior means for rates of deterioration across three diagnosis levels for ventricle and hippocampus volume, credible intervals for estimates in parenthesis. Bottom: Group differences among the three diagnosis levels.

		Regions: units ICV volume/StndAge $\times 10^{-2}$				
	Parameter	Ventricle	Hippocampus			
HC	β_3	0.56 (0.43, 0.63)	-1.3 (-1.8, -0.094)			
MCI	$\beta_3 + \beta_4$	0.66 (0.46, 0.88)	-2.1 (-2.9, -1.4)			
AD	$\beta_3 + \beta_5$	0.96 (0.57, 1.2)	-2.4 (-3.7, -1.3)			
	Est	ge among diagnosis groups				
HC - MCI	β_4	0.094 (-0.072, 0.23)	-0.81 (-1.7, -0.078)			
HC - AD	β_5	0.38 (0.079, 0.70) -1.4 (-2.5, 0.0011)				

Table 2: Posterior probabilities showing comparisons between HC (healthy control), MCI (mild cognitive impaired) and AD (Alzheimer's disease) for ventricle and hippocampus volume. These results provide strong evidence regarding the order of diagnosis levels, derived from the case study.

BMJ Open: first published as 10.1136/bmjopen-2016-012174 on 7 February 2017. Downloaded from http://bmjopen.bmj.com/ on June 13, 2025 at Agence Bibliographique de Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

	0.980	0.991
Hippocampus	$P(AD < MCI) = P(\beta_5 < \beta_4)$	$P(MCI < HC) = P(\beta_5 < 0)$
	0.685	0.985

Our BLME models also allow for probability statements to be made, based on whether any of the slopes are greater or smaller than a biologically meaningful constant or threshold. Table 2 shows the posterior probabilities of deterioration ordering for the three diagnosis categories for ventricle and hippocampus volume, as computed in Eq. (4). The large probabilities support the sequential pattern of deterioration for both regions.

Group comparisons are generally done via hypothesis tests in a classical LME and do not allow for probability statements of group ordering or provide visualization on the distribution of the three groups which quantify the variability in the varying group sizes. The results presented in Table 1 support our hypothesis test results (full analysis in Supplementary material), which show MCI and AD slope were significantly different from baseline for the hippocampus model, whereas only AD slope was significant for the ventricle model. While both classical and Bayesian results can be compared in Table 1, Table 2 and the bottom of Figure 2 can only be produced under the Bayesian framework.

4.2 How to identify individuals with high levels of neurodegeneration?

The rate of deterioration (as measured by the rate of change with respect to age) for the ventricle and hippocampus are in reverse order; large positive ventricle slopes denote high atrophy whereas low negative slope denote large hippocampus atrophy. Table 3 shows a snippet of the participants ranked in order of their estimated median posterior deterioration rate. The data available in this study are highly unbalanced; nonetheless all individuals are ranked despite 19 patients being observed at a single time point only. This is due to the "borrowing strength" aspect of mixed effects models, in that information across all time points contribute to the estimation of the population trends.

Fig. 4 shows clusterings based on HC, MCI and AD participants, denoted by the blue, purple and red box plots, respectively. This reflects the general order of diagnosis rates of deterioration for the ventricle and hippocampus volumes as shown in Fig. 2. However, there are a few individuals who do not follow this pattern, namely those in the small clustered group with the positive estimated levels of atrophy in the hippocampus model and participant ID 1122 in the ventricle model. Participant ID's 1122 and 483 are two out of the 19 individuals who only had baseline measurements, so the rate of deterioration was not observed, but it was still estimated. The same analysis was conducted with a classical LME model and Figure 4 and Table 3 were replicated, refer to the Supplementary material. We found strong similarities with the ranking of the eight converters of interest on both hippocampus and ventricle models.

Table 3: Ranking of individuals from largest to smallest in order of posterior expected rate of deterioration $(\beta_{3i}, \beta_{3i} + \beta_{4i}, \beta_{3i} + \beta_{5i})$ slope for all 260 participants, with 95% credible intervals in parenthesis. Snippet of table shows first and last five individuals, for ventricle and hippocampus volumes. Diagnosis levels; HC, MCI, AD and Converter (either from HC to MCI, HC to AD or MCI to AD), to identify the 27 individuals who changed diagnosis throughout the study, as seen in Fig. 4.

	Ranking	AIBL.ID	Diagnosis	Posterior mean rate of deterioration for individuals $(credible interpole) \times 10^{-2}$
Ventricle	1	1122	AD	-0.14 (-1.1, 0.98)
	2	68	HC	0.23 (-0.19, 0.64)
	3	771	HC	0.28 (-0.13, 0.66)
	4	814	HC	0.29 (-0.054, 0.064)
	5	698	HC	0.29 (-0.011, 0.60)
	256	1032	AD	1.6 (0.55, 2.7)
	257	102	AD	1.6 (0.81, 2.4)
	258	10	AD	1.7 (0.78, 2.7)
	259	658	AD	1.7 (0.94, 2.5)
	260	1102	AD	2.3 (1.5, 3.2)

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Hippocampus	1	10	AD	-9.3 (-16, -3.7)
	2	12	AD	-6.1 (-9.9, -2.4)
	3	1135	AD	-5.7 (-1.1 - 1.0)
	4	398	AD	-5.6 (-9.5, -1.7)
	5	19	AD	-5.4 (-9.8, -1.7)
	256	156	HC	0.02 (-1.5, 1.8)
	257	62	HC	0.41 (-0.95, 1.8)
	258	80	HC	1.0 (-0.54, 2.7)
	259	483	MCI	1.2 (-1.9, 4.6)
	260	1122	AD	1.5 (-1.5, 4.9)

There were 27 individuals who progressed from baseline to a worse diagnosis. Eight individuals of interest are those who progressed from HC to MCI and who had at least three repeated observations recorded. Their estimated deterioration rankings are shown in Fig. 4. The majority of the eight converters in the hippocampus model are scattered along the lower half of the ranking of deterioration. This suggests that their linear rates of hippocampus neurodegeneration are less than those of the AD patients. However patient IDs 757, 232 and 471 were ranked approximately mid-way in this analysis, suggesting that they are approaching hippocampus rates of deterioration similar to AD, and out of the eight converters, they are the most at risk.

Likewise for the ventricle model at the top of Fig. 4, patient ID 471 shows a ventricle rate of deterioration strongly similar to the AD cohort. Further investigation of patient ID 471, such as past family mental history of other forms of dementia, stroke or other mental illness, current cognitive status and other health related factors may provide further insight as to why this individual has an unusually high rate of ventricle deterioration in comparison with the rest of the HC to MCI converters.

*** Figure 3 ***

Fig. 3 shows the posterior distribution of ranks for participants IDs 721 and 12 who converted from MCI to AD at time point 4. Probabilities of these individuals ranked in the lowest 15^{th} quantile for the ventricle volume is 0.75 and 0.46 respectively for participants IDs 721 and 12, likewise for the hippocampus region, these probabilities are 0.47 and 0.58. This same analysis can be performed on any quantile range for any participants of interest. These probabilities show that these participants are in the high neurodegeneration extreme. This same analyses on the full data (over four time points) result in probabilities of participants IDs 721 and 12 ranked in the top 15^{th} quantile are 0.80 and 0.66 for the ventricle and 0.54 and 0.69 for the hippocampus regions. Refer to Supplementary material for posterior ranks distribution plots for all 27 converters. Under the classical implementation of model (3) the distribution of ranks for participants cannot be derived. Once participant ranking is estimated, no probability statements can be made to further analyse individuals at the high or low ranking extremes, and compare for example the high and low 15^{th} quantile extremes. Refer to the Supplementary material for the classical to further analyse individuals at the high or low ranking extremes, and compare for example the high and low 15^{th} quantile extremes. Refer to the Supplementary material for the classical model results.

*** Figure 4 ***

4.3 How do diagnosis trajectories vary over age?

The aim of these analyses is to show the relationship between a volume percentile, combined with results from external sources, to predict region specific diagnosis changes over time. As described in Section 3.5, here we present the probability of a new individual diagnosed as either HC, MCI or AD conditional on volume range and specified age between 60-85 years.

Volume ranges, \tilde{y} , were the 75 – 100^{th} , 50 – 75th, 25 – 50^{th} and $15 - 25^{th}$ percentiles, as shown in Fig. 1 in Section 3.5. Eq. (5) established relationships $P(HC|\tilde{y}, age)$, $P(MCI|\tilde{y}, age)$ and $P(AD|\tilde{y}, age)$ which consists of the output from BLME model stated in (3) in conjunction with prevalence rates from Ward et al. (2012)⁵⁴ and Refshauge and Kalisch (2012)⁵⁵.

*** Figure 5 ***

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Uncertainty in the convergence trajectories of diagnosis levels is presented in terms of probabilities, hence no credible intervals can be estimated. However, there is Monte Carlo error associated with these estimates as they are derived from a finite sample from the posterior distribution. The ventricle and hippocampus models in expression (3) were estimated independently B = 10 times, hence every computation to derive the probability trajectories in this analysis was also estimated ten times in order to compute the Monte Carlo standard error estimates. Let the estimated quantity be denoted as θ and sd be the standard deviation, then a 95% interval for the Monte Carlo standard error is estimated as $\hat{p} \pm 1.96 \times sd(\theta_1, \theta_2, ..., \theta_B)\sqrt{B}$. As $B \to \infty$, the Monte Carlo standard error tends to zero, and while practically B must be finite, our narrow confidence intervals in Fig. 3.5 suggests our simulation methods are adequate for our application.

The results in Fig. 5 show a large difference between HC in contrast with MCI and AD diagnosis for ages 60 to 75 across all ventricle volume quantiles. From age 75 onwards, those individuals in the top percentile range $(75 - 100^{th})$ show the quickest convergence of all the diagnosis levels, who by age 85, show an approximate equal probability (0.30 and 0.31) of being diagnosed as MCI or AD and only a slightly higher chance (0.39) of remaining HC. This contrasts those participants in the lower ventricle volume range $(15 - 25^{th})$, whose difference in diagnosis is vastly different towards the later ages. By age 85, there is a mean estimated 0.60 probability of remaining HC, 0.27 probability of being classified as MCI and an approximate 0.13 probability of AD diagnosis.

The hippocampus model results for this analysis are shown on the bottom of Fig. 5. Between the ages 60-70 there is very little difference across the diagnosis patterns, suggesting individuals whose hippocampus volume lie above the 15^{th} percentile have an approximately equal risk of HC, MCI or AD diagnosis. From age 70 onwards a noticeable difference in diagnosis trajectories is seen across all volume regions, 5 years earlier than the ventricle volume results. This is supported by a large body of literature^{5,33,59-62} as the hippocampus is affected at early stage of development of AD compared to other brain regions. As low hippocampus volume denotes high atrophy, individuals who fall in the lower range volumes, $15 - 25^{th}$ percentile, are most at risk of proceeding onto AD. Individuals in the lower hippocampus volume range, at age 85 have an approximate equal chance of HC, MCI or AD, as shown in Fig. 5.

Diagnosis trajectories over groups; male, female, apolipoprotein $\varepsilon 4$ (APOE $\varepsilon 4$) carriers and non-carriers were also investigated for the hippocampus and ventricle regions utilising model (3). We assumed the same prevalence rates within the population, for example P(MCI|age) = P(MCI|age, female), hence the same broad prevalence rates from Ward et al. (2012)⁵⁴ and Refshauge and Kalisch (2012)⁵⁵ were used. Very little difference in probable disease trajectory across all groups between ages 60 to 85 were observed (refer to the Supplementary material for plots). APOE $\varepsilon 4$ has been associated to an increased likelihood of developing AD.^{63–65} Gender differences regarding the prevalence of AD have also been studied.^{66,67} As the BLME models and inference derivation presented in this paper are the first of their kind, the objective of this analysis is to demonstrate probable diagnosis trajectories conditional on very broad, non-group specific prevalence rates. Future models which account for APOE- $\varepsilon 4$, gender and other factors will utilise group-specific prevalence rates. However to derive the same inference, this would require group specific prevalence rates across ages 60-85, which are difficult to attain from literature.

Our results support those presented in Holland et al. (2012)¹⁵, whereby diagnosis trajectories for neurodegenerated individuals (that is those with very low hippocampus and high ventricle volume) converge at the highest age group, in general over the age of 85. In particular, our results support those presented in Jack et al. (2014)³² for probabilistic trajectory of beta amyloid negative and neurodegeneration positive participants. To make our results comparable to those from Jack et al. (2014)³², HC participants whose hippocampus volume is less than the 50th percentile are defined as neurodegeneration positive. While both methods present trajectories for neurogedegenration of participants over age, the BLME models presented here primarily estimate the rate of volumetric change for the ventricle and hippocampus regions. There are many other inferences that can be deduced from a combination of tapping into the vast wealth of AD research,^{5,68} coupled with the present study analysis. The results presented here are some of the advantages of modelling neurodegeneration through mixed effects models in the Bayesian framework.

Discussion

In this research, we extended the level of insight commonly derived by LME models applied to longitudinal neuroimaging data into three key areas based on a BLME model on the ventricle and hippocampus ICV normalised For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

volumes. We propose that a Bayesian approach for longitudinal neuroimaging modeling has merit for providing further understanding of brain atrophy over time. These views were demonstrated using an application of BLME models applied to a longitudinal AIBL study, which were compared to the classical alternative of LME models.

Comparisons of volumetric rate of change of diagnosis level trajectories were compared for HC, MCI and AD participants, with an estimated probability greater than 0.65 on the order of disease pathology, while credible intervals for the parameters support results from the hypothesis test on a classical LME, under this framework the probability of disease pathology order are not straightforward to compute. Ranking of converters with respect to the study cohort and diagnosis trajectories over age based on volumetric quantiles are the first BLME analysis of their kind applied to longitudinal neuroimaging data. This analysis identified HC to MCI converters most at risk of AD like rate of deterioration and posterior rank distributions provided probabilities on individuals of interest in the worst 15th percentile rank for both regions. The predictive capability of future converters can be derived from these BLME models, as individuals with high neurodegeneration estimates would rank at the extremes in comparison with the remainder of the cohort. The uncertainty of their rank values among a specified quantile is expressed in terms of probabilities, and individuals with a high probability of ranking at extreme levels of neurodegeneration may be indicative of their progressive pathway to further stages of dementia. While classical methods were also able to rank participants in order of estimated volumetric rate of change, they do not allow for further estimation of the highest ranked individuals and the uncertainty in their position. However to rigorously validate this analysis a richer data set with more repeated measures and converters over all categories (HC to MCI or AD and MCI to AD) observed at various ages is required. Furthermore, the diagnosis trajectories for each volume region identified critical points in time for both ventricle and hippocampus degeneration from which participants are most likely to show greater deterioration rates. Alternatively a similar analysis can also be used to compare diagnosis trajectories of different longitudinal neuroimgaing population studies, such as the Alzheimer's Disease Neuroimaging Initiative (ADNI).

Additional analysis regarding group comparisons can be made. For example, similar probabilities for an estimated population mean in comparison to a biologically meaningful constant could also be inferred. An extension to our second analysis to allow population studies to focus on specific participants of interest and monitor their progression rate throughout follow-ups could assist health professionals in making informed choices with regard to patient care. Alternatively HC to AD converters may also be further analysed and ranked with respect to the cohort, to provide further clues as to why these individuals deteriorated so quickly compared with their slower converter counterparts. It is worth- while to note that these inference extensions would not have been possible had we not first attempted the research methods presented in this paper.

A sensitivity analysis with respect to the prior information used in our analysis was conducted on both the ventricle and hippocampus models. This entailed re-running the MCMC sampling technique for each model based on various specifications of the prior information. The subsequent posterior summaries did not vary considerably based on different prior information. Hence, we conjecture that the results are relatively robust to the priors specified in this work. Furthermore, two LOOCV methods were applied to assess the models predictive capability.

Despite every precaution taken to provide robust and reliable conclusions from the BLME models, several authors^{7,14} have noted the limitations and disadvantages of Bayesian statistics applied to longitudinal neuroimaging analysis. In particular, drawbacks of Bayesian statistics in the neuroimaging context are discussed at length in Grunkemeier and Payne (2002)⁶⁹. These include subjective information that can be incorporated in the BLME model specification, in the way in which the prior is specified. Moreover, computational intensity is often far greater in the Bayesian framework than numerical methods employed in a frequentist analysis. In this paper we incorporated vague priors which are semi-conjugate, as we assumed no prior knowledge of the study analysed; the prior specification were a standard choice as suggested in Gelman and Hill (2006)²³. The additional computational insight given. We suggest more complex models and future extensions to the methods presented in this paper may result in an increase of computational time, and this will be a factor to consider for future BMLE models.

Extensions to the BLME models presented in this paper include the addition of more covariates to account for trends and variability sources present in gender, genetic factors and additional demographic characteristics which are a few of the key factors known to affect AD onset and disease progression. Furthermore, as the Bayesian framework is ideal for handling complex models such as generalised linear mixed models^{43,70} and spatio-temporal interactions,^{71,72} extensions of this nature will allow for modelling biomarker deterioration rates of multiple brain regions simultaneously over time.

Acknowledgements We wish to thank the Australian Imaging, Biomarkers and Lifestyle Flagship Study of Ageing (<u>http://www.aibl.csiro.au</u>), including all clinicians, scientist, participants and their families.

Contributorship statement: KM and MIC conceived and designed research concept, CCD provided additional suggestions. Statistical analysis and manuscript drafting was performed by MIC. JD and JF were responsible for the acquisition and interpretation of the data. MIC, JF, JMM, CCD, KM and JD participated in critical revision of the manuscript and approved the final manuscript. MIC is responsible for the overall content as the corresponding author.

Competing interests: The authors declare that they have no competing interests.

Funding: This research was jointly funded by an Australian Postgraduate Award (APA), Commonwealth Scientific and Industrial Research Organisation (CSIRO) Digital Productivity and Services Division, and the ARC Centre of Excellence for Mathematical & Statistical Frontiers.

Data sharing statement: Data available from the Australian Imaging, Biomarkers and Lifestyle longitudinal study of ageing (AIBL). This study is funded by the CSIRO and partners. Access to the data is conditional on

approval from the AIBL management committee, for guidelines refer to http://aibl.csiro.au/research/support/. All R code and simulated data for this manuscript is available at https://github.com/MarcelaCespedes/Bayesian_inference_on_neuroimaging.

1	4	

REFERENCES

- 1. Villemagne, V. L. *et al.* Amyloid β deposition, neurodegeneration and cognitive decline in sporadic Alzheimer's Disease. *Lancet Neurol.* **12**, 357–367 (2013).
- 2. Stoessl, A. J. Neuroimaging in the early diagnosis of neurodegenerative disease. *Transl. Neurodegener.* **1**, (2012).
- 3. Adaszewski, A., Dukart, J., Kherif, F., Frackowiak, R. & Draganski, B. How early can we predict Alzheimer's disease using computational anatomy. *Neurobiol. Aging* **34**, 2815–2826 (2013).
- 4. Mattila, J. *et al.* A disease state fingerprint for evaluation of Alzheimer's disease. *J. Alzheimer's Dis.*27, 163–176 (2011).
- 5. Weiner, M. W. *et al.* The Alzheimer's Disease Neuroimaging Initiative: a review of papers published since its inception. *Alzheimer's Dement.* **9**, e111–e194 (2013).
- Wang, Y., Resnick, S. M., Davatzikos, C., the Baltimore Longitudinal Study of Aging & the Alzheimer's Disease Neuroimaging Initiative. Analysis of spatio-temporal brain imaging patterns by Hidden Markov models and serial MRI images. *Hum. Brain Mapp.* 35, 4777–4794 (2014).
- Bernal-Rusiel, J., Greve, D. N., Reuter, M., Fischl, B. & Sabuncu, M. R. Statistical analysis of longitudinal neuroimage data with Linear Mixed Effects models. *Neuroimage* 66, 249–260 (2013).
- 8. Ziegler, G., Penny, W. D., Ridgway, G. R., Ourselin, S. & Friston, K. J. Estimating anatomical trajectories with Bayesian mixed-effects modeling. *Neuroimage* **121**, 51–68 (2015).
- 9. Khedher, L., Ramírez, J., Górriz, J. M., Brahim, A. & Segovia, F. Early diagnosis of Alzheimer's disease based on partial least squares, principal component analysis and support vector machine using segmented MRI images. *Neurocomputing* **151**, 139–150 (2015).
- 10. Dai, Z. *et al.* Discriminative analysis of early Alzheimer's disease using multi-modal imaging and multi-level characterization with multi-classifier (M3). *Neuroimage* **59**, 2187–2195 (2012).
- 11. Arco, J. E., Ramirez, J., Puntonet, C. G., Gorriz, J. M. & Ruz, M. Improving short-term prediction from MCI to AD by applying searchlight analysis. in *IEEE 13th International Symposium on Biomedical Imaging (ISBI)* 10–13 (2016).
- 12. Arco, J. E., Ramírez, J., Górriz, J. M., Puntonet, C. G. & Ruz, M. in *Innovation in Medicine and Healthcare* 385 – 394 (Springer International Publishing, 2015).
- 13. Ortiz, A., Munilla, J., Illan, I. A., Gorriz, J. M. & Ramírez, J. Exploratory graphical models of functional and structural connectivity patterns for Alzheimer's Disease diagnosis. *Front. Comput. Neurosci.* **9**, 1–18 (2015).
- 14. Guillaume, B., Hua, X., Thompson, P. M., Waldorp, L. & Nichols, T. E. Fast and accurate modelling of longitudinal and repeated measures neuroimaging data. *Neuroimage* **94**, 287–302 (2014).
- 15. Holland, D. *et al.* Rates of decline in Alzheimer disease decrease with age. *PLoS One* **7**, e42325 (2012).
- 16. Kruschke, J. K. Bayesian estimation supersedes the t test. J. Exp. Psychol. Gen. 142, 573 (2013).
- Pfefferbaum, A. *et al.* Variation in longitudinal trajectories of regional brain volumes of healthy men and women (ages 10 to 85years) measured with atlas-based parcellation of MRI. *Neuroimage* 65, 176–193 (2013).
- 18. Jack Jr, C. R. *et al.* Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. *Lancet {N}eurology* **12**, 207–216 (2013).

- 19. Oxtoby, N. P. *et al.* in *Bayesian and grAphical Models for Biomedical Imaging* 85–94 (Springer, 2014).
- 20. Sperling, R. A. *et al.* Toward defining the preclinical stages of Alzheimer's disease: Recommnedations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's Dement.* **7**, 280–292 (2011).
- 21. Ferreira, L. K. & Busatto, G. F. Neuroimaging in Alzheimer's disease: current role in clinical practice and potential future applications. *Clinics* **66**, 19–24 (2011).
- 22. Harville, D. A. & Carriquiry, A. L. Classical and Bayesian prediction as applied to an unbalanced mixed linear model. *Biometrics* 987–1003 (1992).
- 23. Gelman, A. & Hill, J. *Data analysis using regression and multilevel/hierarchical models*. (Cambridge University Press, 2006).
- 24. Li, R., Stewart, B. & Weiskittel, A. A Bayesian approach for modelling non-linear longitudinal/hierarchical data with random effects in forestry. *Forestry* **85**, 17–25 (2012).
- 25. Friston, K. J. *et al.* Classical and Bayesian inference in neuroimaging: Theory. *Neuroimage* **16**, 465–483 (2002).
- 26. Scheltens, P. Advances in Neuroimaging. *Blue Books Neurol.* **30**, 381–410 (2007).
- 27. Risacher, S. L. & Saykin, A. J. Neuroimaging and other biomarkers for {A}lzheimer's disease: the changing landscape of early detection. *Annu. Rev. Clin. Psychol.* **9**, 621–622 (2013).
- 28. Rodie, M. E., Forbes, K. P. & Muir, K. in *Understanding Differences and Disorders of Sex Development* (*DSD*) (eds. Hiort, O. & Ahmed, S. F.) **27**, 63–75 (Karger Publishers, 2014).
- 29. Jack Jr, C. R. *et al.* Tracking pathophysiological processes in {A}lzheimer's disease: an updated hypothetical model of dynamic biomarkers. *Lancet* {*Neurology* **12**, 207–216 (2013).
- 30. Ziegler, G., Penny, W. D., Ridgway, G. R., Ourselin, S. & Friston, K. J. Estimating anatomical trajectories with Bayesian mixed-effects modeling. *Neuroimage* **121**, 51–68 (2015).
- 31. Sperling, R. A., Karlawish, J. & Johnson, K. A. Preclinical Alzheimer disease the challenges ahead. *Nat. Rev. Neurol.* **9**, 54–58 (2013).
- Jack, C. R. *et al.* Age-specific population frequencies of cerebral β-amyloidosis and neurodegeneration among people with normal cognitive function aged 50--89 years: a crosssectional study. *Lancet Neurol.* 13, 997–1005 (2014).
- 33. Apostolova, L. G. *et al.* Hippocampal atrophy and ventricular enlargement in normal aging, mild cognitive impairment and Alzheimer's disease. *Alzheimer Dis. Assoc. Disord.* **26**, 17 (2012).
- 34. Tang, X., Holland, D., Dale, A. M., Younes, L. & Miller, M. I. Shape abnormalities of subcortical and ventricular structures in mild cognitive impairment and Alzheimer's disease: detecting, quantifying, and predicting. *Hum. Brain Mapp.* (2014).
- 35. Chupin, M. *et al.* Fully automatic hippocampus segmentation and classification in Alzheimer's disease and mild cognitive impairment applied on data from ADNI. *Hippocampus* **19**, 579–587 (2009).
- 36. Jack Jr, C. R. *et al.* Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *Lancet Neurol.* **9**, 119–128 (2010).
- 37. Huijbers, W. *et al.* Amyloid-B deposition in mild cognitive impairment is associated with increased hippocampal activity, atrophy and clinical progression. *Brain* **138**, 1023–1035 (2015).
- 38. Rowe, C. C. *et al.* Amyloid imaging results from the Australian Imaging, Biomarker and Lifestyle For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

7 of 38	BMJ Open
16	
	(AIBL) study of aging. Neurobiol. Ageing 31, 1275–1283 (2010).
39.	O'Brien, L. M. <i>et al.</i> Statistical adjustments for brain size in volumetric neuroimaging studies: Some practical implications in methods. <i>Psychiatry Res. Neuroimaging</i> 193 , 113–122 (2011).
40.	Voevodskaya, O. <i>et al.</i> The effects of intracranial volume adjustment approaches on multiple regional MRI volumes in healthy aging and Alzheimer's disease. <i>Front. Aging Neurosci.</i> 6, (2014).
41.	Pinheiro, J. C. & Bates, D. M. Mixed-effects models in S and S-PLUS. (Springer, 2000).
42.	McCullagh, P. Generalized linear models. <i>Eur. J. Oper. Res.</i> 16, 285–292 (1984).
43.	Gelman, A. et al. Bayesian data analysis. (CRC press, 2013).
44.	Robert, C. & Casella, G. Monte Carlo Statistical Methods. (Springer New York, 2010).
45.	Burden, R. & Faires, J. Numerical analysis. (Cengage Learning, 2011).
46.	Adler, J. R in a nutshell: A desktop quick reference. (' O'Reilly Media, Inc.', 2010).
47.	Plummer, M., Best, N., Cowles, K. & Vines, K. CODA: Convergence Diagnosis and Output Analysis for MCMC. <i>R News</i> 6 , 7–11 (2006).
48.	Wickham, H. <i>ggplot2: elegant graphics for data analysis</i> . (Springer New York, 2009). at <http: book="" ggplot2="" had.co.nz=""></http:>
49.	Gelman, A. & Rubin, D. B. Inference from iterative simulation using multiple sequences. <i>Stat. Sci.</i> 7, 457–472 (1992).
50.	Wang, W. & Gelman, A. Difficulty of selecting among multilevel models using predictive accuracy. <i>Stat. its Interface</i> 7 , 1–15 (2014).
51.	Jack Jr, C. R. <i>et al.</i> Shapes of the trajectories of five major biomarkers of Alzheimer's Disease. <i>Arch.</i> <i>Neurol.</i> 7, 856–867 (2012).
52.	Woolrich, M. W., Behrens, T. E. J. & Smith, S. M. Constrained linear basis sets for HRF modelling using Variational Bayes. <i>Neuroimage</i> 21, 1748–1761 (2004).
53.	Bernal-Rusiel, J. L., Reuter, M., Greve, D. N., Fischl, B. & Sabuncu, M. R. Spatiotemporal linear mixed effects modeling for the mass-univariate analysis of longitudinal neuroimage data. <i>Neuroimage</i> 81 , 358–370 (2013).
54.	Ward, A., Arrighi, H. M., Michels, S. & Cedarbaum, J. M. Mild cognitive impairment: disparity of incidence and prevalence estimates. <i>Alzheimer's Dement.</i> 8, 14–21 (2012).
55.	Refshauge, A. & Kalisch, D. Dementia in Australia. <i>Australian Institute of Health and Welfare</i> (2012). at <http: downloadasset.aspx?id="10737422943" workarea="" www.aihw.gov.au=""></http:>
56.	Cole, T. J. & Green, P. J. Smoothing reference centile curves: the LMS method and penalized likelihood. <i>Stat. Med.</i> 11, 1305–1319 (1992).
57.	Koenker, R. Quantile regression. (Cambridge university press, 2005).
58.	Muggeo, V. M. R., Sciandra, M., Tomasello, A. & Calvo, S. Estimating growth charts via nonparametric quantile regression: a practical framework with application in ecology. <i>Environ. Ecol. Stat.</i> 20 , 519–531 (2013).
59.	Caroli, A. & Frisoni, G. B. The dynamics of Alzheimer's disease biomarkers in the Alzheimer's Disease Neuroimaging cohort. <i>Neurobiol. Aging</i> 31, 1263–1274 (2010).
60.	Leung, K. K. <i>et al.</i> Cerebral atrophy in mild cognitive impairment and Alzheimer's disease. <i>Am. Acad. Neurol.</i> 80, 648–654 (2013).

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

- 61. Risacher, S. L. & Saykin, A. J. Neuroimaging and other biomarkers for Alzheimer's disease: the changing landscape of early detection. *Annu. Rev. Clin. Psychol.* **9**, 621–622 (2013).
- 62. Antoniano-Villalobos, I., Wade, S. & Walker, S. G. A Bayesian nonparametric regression model with normalized weights: A study of hippocampal atrophy in Alzheimer's disease. *J. Am. Stat. Assoc.* **109**, 477–490 (2014).
- 63. Caselli, R. J. *et al.* Longitudinal modeling of age-related memory decline and the APOE e4 effect. *N. Engl. J. Med.* **361**, 255–263 (2009).
- 64. Verghese, P. B., Castellano, J. M. & Holtzman, D. M. Apolipoprotein E in Alzheimer's disease and other neurological disorders. *Lancet Neurol.* **10**, 241–252 (2011).
- 65. Liu, C. C., Kanekiyo, T., Xu, H. & Bu, G. Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy. *Nat. Rev. Neurol.* **9**, 106–118 (2013).
- 66. Carter, C. L., Resnick, E. M., Mallampalli, M. & Kalbarczyk, A. Sex and gender differences in Alzheimer's disease: recommendations for future research. *J. Women's Heal.* **21,** 1018–1023 (2012).
- 67. Li, R. & Singh, M. Sex differences in cognitive impairment and Alzheimer's disease. *Front. Neuroendocrinol.* **35**, 385–403 (2014).
- 68. Hardy, J. A hundred years of Alzheimer's disease research. *Neuron* **52**, 3–13 (2006).
- 69. Grunkemeier, G. L. & Payne, N. Bayesian analysis: a new statistical paradigm for new technology. *Ann. Thorac. Surg.* **74**, 1901–1908 (2002).
- 70. Kleinman, K. P. & Ibrahim, J. G. A semi-parametric Bayesian approach to generalized linear mixed models. *Stat. Med.* **17**, 2579–2596 (1998).
- 71. Alston, C. L., Mengersen, K. L. & Pettitt, A. N. *Case studies in Bayesian statistical modelling and analysis*. (John Wiley & Sons, 2012).
- 72. Banerjee, S., Carlin, B. P. & Gelfand, A. E. *Hierarchical modeling and analysis for spatial data*. (Crc Press, 2014).

BMJ Open

Figure legends

Fig. 1: Percentile ranges of volume across ages 60 to 85 years old, for ventricle (left) and hippocampus (right). Recall region volumes are normalised by the ICV value as they represent a percentage of volume within the intracranial cavity. Ranges up to the 100^{th} percentile henceforth denote the empirical maximum volume for that region. Volume percentiles; $75 - 100^{th}$ from blue (0.75) to top dotted line, $50 - 75^{th}$ from green (0.25) to blue (0.75) line, $25 - 50^{th}$ from red (0.25) to green (0.50) line and $15 - 25^{th}$ from black (0.15) to red (0.25) line.

Fig. 2: Posterior densities of population mean estimates of linear deterioration rate for diagnosis (top plot): HC, MCI and AD, for ventricle (left) and hippocampus volume (right) models. Dotted lines on bottom plots denote the means for each density, whose values are shown in Table 1.

Fig. 3: Posterior distribution of ranks for MCI to AD converters ID 721, 365 and 12, for Ventricle (top) and hippocampus (bottom) ICV volume models. These density rankings were derived with observations from time points one to three.

Fig. 4: Box plots of posterior distribution of random effect values for participants in AIBL study (N = 260) for full data (four time points). Ventricle (top) and hippocampus (bottom) rates of deterioration for each participant in the study. As there are 157 HC, 34 MCI, 42 AD and 27 Converters in this study, there is higher uncertainty on the rate of deterioration of converters, MCI and AD participants (hence longer box plots) as compared to the HC (narrower box plots). Eight individuals who converted from HC to MCI throughout the study are highlighted in red with corresponding ID numbers.

Fig. 5: Probability curves show the posterior probability of HC, MCI or AD diagnosis for the ventricle (top) and hippocampus (bottom) models, 95% interval denotes Monte Carlo error based on several simulations of the BLME models. Total volume divided into four percentile volume ranges, as shown in Fig. 1. Percentiles; $75 - 100^{th}$, $50 - 75^{th}$, $25 - 50^{th}$ and $15 - 25^{th}$.

BMJ Open: first published as 10.1136/bmjopen-2016-012174 on 7 February 2017. Downloaded from http://bmjopen.bmj.com/ on June 13, 2025 at Agence Bibliographique de l Enseignement Superieur (ABES)

Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies





Fig. 1: Percentile ranges of volume across ages 60 to 85 years old, for ventricle (left) and hippocampus (right). Recall region volumes are normalised by the ICV value as they represent a percentage of volume within the intracranial cavity. Ranges up to the 100^th percentile henceforth denote the empirical maximum volume for that region. Volume percentiles; 75-100^th from blue (0.75) to top dotted line, 50-75^th from green (0.25) to blue (0.75) line, 25- 50^th from red (0.25) to green (0.50) line and 15-25^th from black (0.15) to red (0.25) line.

112x55mm (300 x 300 DPI)





Fig. 2: Posterior densities of population mean estimates of linear deterioration rate for diagnosis (top plot): HC, MCI and AD, for ventricle (left) and hippocampus volume (right) models. Dotted lines on bottom plots denote the means for each density, whose values are shown in Table 1.

173x193mm (300 x 300 DPI)

Page 22 of 38

BMJ Open: first published as 10.1136/bmjopen-2016-012174 on 7 February 2017. Downloaded from http://bmjopen.bmj.com/ on June 13, 2025 at Agence Bibliographique de I Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

BMJ Open



Fig. 3: Posterior distribution of ranks for MCI to AD converters ID 721, 365 and 12, for Ventricle (top) and hippocampus (bottom) ICV volume models. These density rankings were derived with observations from time points one to three.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml



Fig. 4: Box plots of posterior distribution of random effect values for participants in AIBL study (N = 260) for full data (four time points). Ventricle (top) and hippocampus (bottom) rates of deterioration for each participant in the study. As there are 157 HC, 34 MCI, 42 AD and 27 Converters in this study, there is higher uncertainty on the rate of deterioration of converters, MCI and AD participants (hence longer box plots) as compared to the HC (narrower box plots). Eight individuals who converted from HC to MCI throughout the study are highlighted in red with corresponding ID numbers.

158x165mm (300 x 300 DPI)



Fig. 5: Probability curves show the posterior probability of HC, MCI or AD diagnosis for the ventricle (top) and hippocampus (bottom) models, 95% interval denotes Monte Carlo error based on several simulations of the BLME models. Total volume divided into four percentile volume ranges, as shown in Fig. 1. Percentiles; 75- 100th, 50-75th, 25- 50th and 15-25th.

155x345mm (300 x 300 DPI)

Supplementary Material

Please refer to website

https://github.com/MarcelaCespedes/Bayesian_inference_on_neuroimaging for R code (including code for plots) used in analysis outlined in the manuscript *Comparisons of neurodegeneration* over time between healthy ageing and Alzheimer's disease cohorts via Bayesian inference.

1 Derivation of model

In this section, we compare competing models for neurodegeneration with respect to age. Competing models include; linear, quadratic, cubir and quartic configurations, see equations (1) to (4), where (.) denotes the the linear predictor in model (1).

 $Y_{ij}|\mu_{ij},\sigma^2 \sim N(\mu_{ij},\sigma^2)$

 $\mu_{ij} = \beta_{0i} + \beta_1 x_{MCI,ij} + \beta_2 x_{AD,ij} + \beta_{3i} StndAge_{ij} + \beta_{4i} StndAge_{ij} x_{MCI,ij} + \beta_{5i} StndAge_{ij} x_{AD,ij}$ $\beta_{ki} = \beta_k + b_{ki} \quad \text{for } k = 0, 3, 4, 5$ $\boldsymbol{b_i} \sim MVN(\boldsymbol{0}, \Sigma) \tag{1}$

$$\mu_{ij} = (.) + \beta_6 StndAge_{ij}^2 + \beta_7 StndAge_{ij}^2 x_{MCI,ij} + \beta_8 StndAge_{ij}^2 x_{AD,ij}$$

$$\mu_{ij} = (.) + \beta_6 StndAge_{ij}^2 + \beta_7 StndAge_{ij}^2 x_{MCI,ij} + \beta_8 StndAge_{ij}^2 x_{AD,ij} + \beta_9 StndAge_{ij}^3 + \beta_{10} StndAge_{ij}^3 + \beta_{11} StndAge_{ij}^3 x_{AD,ij}$$
(3)

$$\begin{split} \mu_{ij} &= (.) + \beta_6 StndAge_{ij}^2 + \beta_7 StndAge_{ij}^2 x_{MCI,ij} + \beta_8 StndAge_{ij}^2 x_{AD,ij} + \\ \beta_9 StndAge_{ij}^3 + \beta_{10} StndAge_{ij}^3 + \beta_{11} StndAge_{ij}^3 x_{AD,ij} + \\ \beta_{12} StndAge^4 + \beta_{13} StndAge^4 x_{MCI,ij} + \beta_{14} StndAge_{ij}^4 x_{AD,ij} \end{split}$$

(4)

In Bayesian statistics, model choice can be handled via the posterior model probabilities. These probabilities can be estimated straightforwardly via normalising the model evidences, and, as the model evidence provides an inbuilt penalty for model complexity, there is a preference for the model with the largest value (MacKay, 2003). In this work, the integrated nested Laplace approximation (INLA) (Rue et al., 2009) was used to approximate the model evidences, and the logarithm of these values are shown in Table S1.

2 Assessment of normality assumption for models

Below are histograms of the residuals, scatter and quantile-quantile plots for the Ventricle and Hippocampus models presented in the manuscript in expression (3). Refer to Section 3 of the manuscript for assumptions of linear mixed effects models. Figure S1 shows residuals from both models are approximately normal, despite our response values being in (0, 1) range. Refer to Section 2 of the manuscript, for further discussion. BMJ Open: first published as 10.1136/bmjopen-2016-012174 on 7 February 2017. Downloaded from http://bmjopen.bmj.com/ on June 13, 2025 at Agence Bibliographique de Enseignement Superieur (ABES)

BMJ Open: first published as 10.1136/bmjopen-2016-012174 on 7 February 2017. Downloaded from http://bmjopen.bmj.com/ on June 13, 2025 at Agence Bibliographique de Enseignement Superieur (ABES). Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies

60

Model fitted, linear predictor with	Ventricle	Hippocampus
Age(1)	-1231	-1510
$Age + Age^2$ (2)	-1206	-1129
$Age + Age^2 + Age^3$ (3)	-1178	-1101
$Age + Age^2 + Age^3 + Age^4 (4)$	-1150	-1073

Table S1: Log evidence for competing models with non-linear terms with respect to age. We can see that the model with the highest log-evidence for both ventricle and hippocampus models consists of Age as a linear term, which is expression (1) as this has the highest log-evidence values for both regions.



Figure S1: Ventricle (left) and hippocampus (right) assess linear mixed effect model for violation of normality assumptions.

3 Cross validation

We performed cross validation on the hippocampus and ventricle models described in the manuscript, to assess the predictive capability of a new observation. As discussed in Wang and Gelman (2014), there are no clear protocols for cross validation methods for multilevel models and out of sample validation methods for hierarchical models are not as straightforward as a random sample of the data for a holdout set. To that end, we carried out two approaches for leave-one-out cross validation (LOOCV) on the ventricle and hippocampus models presented in the manuscript.

Firstly, all observations for an individual were removed and the model was estimated with the remaining data, and was used to predict the observations for the missing individual. The posterior mean for each predicted value was subtracted from the known observation to attain a residual value. The residuals and predictive means were assessed for the predictive capability of an individual. This was performed on all individuals, including those with single observations and converters.

The second LOOCV technique involved randomly omitting one observation on individuals with repeated measures (199 participants in our data set). As our longitudinal data are unbalanced, this method assessed predicting values across all time points. Predicted values were computed as described above and the residual and predictive values were assessed for predictive capability within clustered groups.



Figure S2: LOOCV-individual, residual (left), residual histogram (middle) and predicted versus response plots (right), for the ventricle (top) and hippocampus (bottom) models.

Left top and bottom plots in Figures S2 and S3 show that the variability in the predictions for out of sample data preserved the overall linear trend and there are no general pattern in the residual scatter plots. As the MSE is the sum of the variance, bias squared and irreducible error, in practise we seek to reduce both bias and variance, hence low MSE values are preferred. Nonetheless there MSE values are relatively low for both models, hence we are satisfied the models stated in (3) in the manuscript do not over-fit the data and provide adequate predictions of new data.

	V	<i>V</i> entricle	Hippocampus		
	LOO-individual	LOO-within-a-cluster	LOO-individual LOO-within-a-cluster		
MSE	$1.86e^{-4}$	$1.42e^{-4}$	$3.91e^{-4}$	$1.66e^{-3}$	

Table S2: Mean squared root error (MSE) for leave-one-out (LOO) on an individuals set of observations and within a cluster cross validation. The posterior mean for predictive value is \tilde{y}_i and observed response is y_i for n observations computed as MSE= $\sum (\tilde{y}_i - y_i)^2/n$ as described in Timm (1980). Note both MSE values are relatively similar for the hippocampus and ventricle models.

BMJ Open: first published as 10.1136/bmjopen-2016-012174 on 7 February 2017. Downloaded from http://bmjopen.bmj.com/ on June 13, 2025 at Agence Bibliographique de Enseignement Superieur (ABES) .

Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies



Figure S3: LOOCV-within-a-cluster, residual (left), residual histogram (middle) and predicted versus response plots (right), for the ventricle (top) and hippocampus (bottom) models.

4 Classical linear mixed effects model

In an effort to compare model (3) in the manuscript with popular conventional longitudinal methods, the model was also fitted by a classical linear mixed effects (LME) model. Recall the general LME model is of the following form,

$$\mathbf{y} = X\boldsymbol{\beta} + Z\mathbf{b} + \varepsilon. \tag{5}$$

Design matrices are X and Z, and vectors β and b are the fixed and random effects respectively for p fixed and m random effects. We assume residuals $\varepsilon \sim N(0, \sigma^2 I)$, where I is the identity matrix. The random effects vector b, assume $\mathbf{b} \sim N(\mathbf{0}, \sigma^2 D(\theta))$, where $D(\theta)$ is a symmetric and positive semi-definite matrix, parametrized by a variance component θ . For further details including methods to estimate the maximisation of the likelihood, see (Pinheiro, 1994; Corbeil and Searle, 1976).

The classical LME model for this case study is denoted as follows

$$Y_{ij} = \beta_{0i} + \beta_{1i}StndAge_{ij} + \beta_{2}x_{MCI} + \beta_{3}x_{AD} + \beta_{4i}StndAge_{ij}x_{MCI,ij} + \beta_{5i}StndAge_{ij}x_{AD,ij} + \varepsilon_{ij}$$

$$\beta_{ki} = \beta_{k} + b_{ki} \quad for \quad k = 0, 1, 4, 5$$

$$\boldsymbol{b_{i}} \sim MVN(\boldsymbol{0}, \Sigma).$$
(6)

Random effects $\mathbf{b}_i = [b_{0i}, b_{1i}, b_{4i}, b_{5i}]$ denotes the i^{th} individual deviation at time point j, and covariance structure Σ is 4×4 diagonal matrix. Residuals ε are i.i.d with distribution $N(0, \sigma^2 I_n)$, for n total observations.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

		Regions: units ICV volume/StndAge					
	Parameter	Ventricle	Hippocampus				
HC	β_3	$5.99e^{-3} (3.29e^{-4})$	$-1.18e^{-2} (1.54e^{-3})$				
MCI	$\beta_3 - \beta_4$	$6.25e^{-3} (6.69e^{-4})$	$-2.10e^2 (5.61e^{-3})$				
AD	$\beta_3 - \beta_4$	$1.10e^{-3} (5.36e^{-3})$	$-2.82e^{-2}$ (7.12 e^{-3})				
	Estimated c	lifference of volumetri	ic change among diagnosis groups				
HC - MCI	β_4	$2.68e^{-3} (3.40e^{-4})$	$-9.27e^{-3} (4.07e^{-3})$				
HC - AD	β_5	$5.03e^{-3} (1.49e^{-3})$	$-1.64e^{-2}(5.59e^{-3})$				

R package lme4() (Bates et al., 2014b), was used to estimate model (6) (R Core Team, 2013).

Table S3: Parameter estimates for model (6) for both regions, standard error in parenthesis. Similar to Table 1 in the manuscript, the fixed effect values are similar to the BLME model.

4.1 How do HC, MCI and AD participants degenerate over time?

Hypothesis tests were conducted in a similar manner to Bernal-Rusiel et al. (2013) for the ventricle and hippocampus models, refer to their supplementary material for their full model expressions, contrast matrices and null hypothesis statements. In a similar manner, we also state our null hypothesis to compare the rate of volumetric rate of change on diagnosis groups HC, MCI and AD for both brain regions, and set the significance level to $\alpha = 0.05$.

Test 1: Are there any differences in the rate of change among the three groups? The null hypothesis is $H_0: \beta_4 = \beta_5 = 0$. The contrast matrix for both the ventricle and hippocampus models are of the form

BMJ Open: first published as 10.1136/bmjopen-2016-012174 on 7 February 2017. Downloaded from http://bmjopen.bmj.com/ on June 13, 2025 at Agence Bibliographique de Enseignement Superieur (ABES)

Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies

$T_1 =$	0	0	0	0	1	$\left[0 \right]$	
	0	0	0	0	0	1	

	Ventricle	Hippocampus
F-statistic	5.850	6.10
p-value	0.0048	0.0058

Table S4: Results for the hypothesis Test 1 for Ventricle and Hippocampus models. As the p-value is less than α for both tests, we reject the null hypothesis and conclude there is a β_4 and β_5 are not equal to zero in both models.

*Test 2:*Is there any difference in the rate of volumetric change between HC and MCI? The null hypothesis is $H_0: \beta_4 = 0$, with a contrast matrix for both regions as

$$T_2 = [0\ 0\ 0\ 0\ 1\ 0]$$

Test 3: Is there a difference in the rate of change between HC and AD? Similar to above the null hypothesis is $H_0: \beta_5 = 0$, with a contrast matrix for both regions as

$$T_3 = [0\ 0\ 0\ 0\ 0\ 1]$$

Z
3
4
5
5
6
7
8
õ
9
10
11
12
12
13
14
15
16
47
17
18
19
20
20
21
22
23
20
24
25
26
27
21
28
29
30
21
31
32
33
34
25
30
36
37
38
20
39
40
41
42
40
43
44
45
16
40
47
48
49
50
50
51
52
53
55
54
55
56
57
57
58
EΟ

1

	Ventricle	Hippocampus
F-statistic	0.623	5.19
p-value	0.431	0.031

Table S5: Results for the hypothesis Test 2 for Ventricle and Hippocampus models. The Ventricle model indicates we have insufficient evidence to reject the null hypothesis, and conclude the rate of change for MCI is not significantly different from baseline. Unlike the hippocampus model, there is sufficient evidence to suggest the rate of change for MCI is significantly different from baseline.

	Ventricle	Hippocampus
F-statistic	11.39	8.65
p-value	0.0018	0.0061

Table S6: Results for the hypothesis Test 3 for Ventricle and Hippocampus models. As both p-values are less than α , we reject the null hypothesis and conclude the rate of change for AD diagnosis is significantly different from baseline.

4.2 How to identify individuals with high levels of neurodegeneration?

The second analysis presented in the manuscript in Sections 3.4 and 4.2 is presented here where possible, to assess individual participants rate of change. The caterpillar plots for each region are in Figure S4a and S4b which were derived from expression (6) above.

Caterpillar plots for classical mixed effects models are often used for analysis of the random effects (Bates et al., 2014a). The uncertainty for each individual is estimated by conditional variances of the random effects from the output of lmer(). The general pattern of HC to MCI converters is similar to those in the manuscript. The ventricle model shows ID's 4, 113, 737 and 509 and the hippocampus participant ID's 757, 232, 471 are in the lower half of the ranks similar to the Bayesian model Figure 4 in the manuscript.

Unfortunately, under this framework we cannot determine the probability of participants ranking in the highest or lowest degeneration extremes. As the ranking of participants relies on point estimates of the random effects and conditional variances, which does not include probability distribution to account for uncertainty between and within observations within clusters.





BMJ Open: first published as 10.1136/bmjopen-2016-012174 on 7 February 2017. Downloaded from http://bmjopen.bmj.com/ on June 13, 2025 at Agence Bibliographique de I Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

Figure S4: Individuals ranked by order of estimated ventricle (a) and hippocampus (b) rate of change.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

BΜ

		~
		g
		en
		=
		S
		5
		B
		Ī
		he
		ã
		S
υ		10
õ		<u>-</u>
e C		36
te		<u>j</u>
<u> </u>		3
<		ဗ
<u></u>		en
2		ż
<u>.</u>		20
7		5
3'		22
<u><u></u></u>		17
2		4
j		ă
Ť		7 F
Ĭ		ē
S	Ш	ž
BS	su	lar
ſe	eic	~
at	<u>b</u>	õ
PO	ž	.7
đ	en.	
te	ົດ	ž
¥	Ę	<u>n</u>
an	ē	ă
<u>0</u>	ē	ē
a	=	ŧ
2	R	<u>o</u>
≝.	Ĕ	2
Ĭ.	<u></u>	₫
ē	•	ä
≥		b
ŧ		믕
a.		ğ
Ž		Ž
2		B
ž		5
s		9
3		2
a		В
2		Ľ
PC.		Ine
ק		-
		ω
ě.		202
PS.		50
•		at
		Å
		Jer
		J CE
		Ē
		D
		Ĩ
		är
		ա
		٩đ
		aphiq

de

	Ranking	AIBL.ID	Diagnosis	Posterior mean rate of deterioration	
	_		_	for individuals (standard error)	
Ventricle	1	877	AD	$-1.04 \times 10^{-2} (8.81 \times 10^{-5})$	
	2	1122	HC	$-8.66 \times 10^{-3} (3.94 \times 10^{-3})$	
	3	911	HC	$-8.45 \times 10^{-3} (1.66 \times 10^{-4})$	
	4	111	HC	$-7.86 \times 10^{-3} (1.20 \times 10^{-2})$	
	5	365	HC	$-5.11 \times 10^{-3} (2.16 \times 10^{-4})$	
	:	:	÷	÷	
	256	658	AD	$1.60 \times 10^{-2} (2.48 \times 10^{-3})$	
	257	1032	AD	$1.61 \times 10^{-2} (5.69 \times 10^{-3})$	
	258	10	AD	$1.67 \times 10^{-2} (\ 7.34 \times 10^{-3})$	
	259	1102 AD 2.3		$2.37 \times 10^{-2} (3.45 \times 10^{-3})$	
	260	91	AD	$2.478 \times 10^{-2} (3.35 \times 10^{-3})$	
Hippocampus	1	10	AD	$-7.42 \times 10^{-2} (2.78 \times 10^{-2})$	
	2	1135	AD	$-6.35 \times 10^{-2} (2.62 \times 10^{-2})$	
	3	12	Converter	$-5.19 \times 10^{-2} (2.34 \times 10^{-2})$	
	4 1013		AD	$-5.13 \times 10^{-2} (2.39 \times 10^{-2})$	
	5	819	AD	$-4.98 \times 10^{-2} (1.86 \times 10^{-2})$	
	:		÷	: :	
	256	483	MCI	$2.42 \times 10^{-2} (1.94 \times 10^{-3})$	
	257	113	Converter	$2.56 \times 10^{-2} (2.55 \times 10^{-3})$	
	258	509	Converter	$2.72 \times 10^{-2} (2.75 \times 10^{-3})$	
	259	28	HC	$3.36 \times 10^{-2} (2.68 \times 10^{-3})$	
	260	1122	AD	$3.71 \times 10^{-2} (1.78 \times 10^{-2})$	

Table S7: Similar to Table 3 in the manuscript, participants by order of estimated rate of change, standard error in parenthesis. Snippet of table shows first and last five individuals for the ventricle and hippocampus volumes.

4.3 How do diagnosis trajectories vary over age?

As described in the manuscript, in this analysis $P(Diagnosis | \tilde{y}, age)$ is estimated for Diagnosis = HC, MCIand AD, as shown in expression (5) in the manuscript. We note that in order to find these probabilities, for a given range in volume \tilde{y} we need the probability of this range given a diagnosis classification and age, i.e. $P(\tilde{y}|Diagnosis, age)$.

A similar analysis cannot be performed with a classical LME model, as the method of maximisation of the likelihood does not allow for the straightforward computation of probabilities $P(Diagnosis|\tilde{y}, age)$. Another drawback of the classical approach is that it does not lend itself to the incorporation of relevant external data, to further extend statistical inference.

5 Posterior Predictive checks and parameter estimates

Posterior predictive checks were carried out to assess goodness-of-fit of our models in expression (3) of the manuscript, as predicted values were simulated from the joint posterior distribution. After burn-in and thinning, as specified in Section 3.2 of the manuscript, each predicted value consists of 8,000 simulations from which we compute the 95% credible intervals. Posterior predictive plots are shown in Figure S5. MCMC chain diagnostics such as trace, density and auto-correlation plots as well as the Gelman and Rubin convergence measures are available





Figure S5: Posterior predictive means versus response values with the 95% credible interval. The tight bandwidth on all responses shows we have adequately captured the variability. As both the plots show a general diagonal pattern of x = y for majority of the values (with the exception of a few cases), this provides evidence of accurate predicted values from our model.

1
2
3
4
5
6
7
8
ğ
10
11
10
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
20
21
28
29
30
31
32
33
34
35
36
37
38
39
40
41
12
72 /2
40
44
40
46
47
48
49
50
51
52
53
54
55
56
57
58
50
59
00

		Ventricle ESS	Hippocampus ESS
	β_{0i}	6895	5625
	β_{1i}	7231	2830
	β_2	5585	3832
	β_3	3463	2779
	β_{4i}	6657	1313
	β_{5i}	5693	544
	σ^2	7662	4810
	σ_0^2	7351	3023
	σ_1^2	7753	1824
	σ_4^2	6670	977
	σ_5^2	5352	1021
	P.P.	0.993	0.995

Table S8: Posterior proportion of response (P.P), is a proportion of predicted values which lie within 95% credible interval of prediction values as seen in Figure S5. Effective sample size (ESS) denotes the estimated number of independent samples (no auto-correlation) obtained in our estimated parameters. As per our burn-in and thinning specifications stated in Section 3.2 of the manuscript, the ESS will be at most a value up to 8,000.

Distribution of ranks for converters 6

As described in Section 4.2 of the manuscript, distribution of ranks were performed on all (27) converters of the AIBL study, first on a subset of the first three time points; for ventricle model see Figure S6, hippocampus see Figure S7. Similarly the distribution ranks were estimated on the whole data set, Figure S8 shows the results for the ventricle model, and Figure S9 correspond to the hippocampus model.



Figure S6: Ventricle converters posterior distribution of ranks for the first three time points.



Figure S7: Hippocampus converters posterior distribution of ranks for the first three time points.



Figure S8: Ventricle converters posterior distribution of ranks for full data (4 timepoints).





Figure S9: Hippocampus converters posterior distribution of ranks for full data (4 timepoints).

APOE and Gender diagnosis trajectories over age 7

As mentioned in Section 4.3 of the manuscript, initial exploration of diagnosis trajectories over groups; male, female, apolipoprotein $\varepsilon 4$ (APOE $\varepsilon 4$) carriers and non-carriers were also investigated for the ventricle and hippocampus models.

The broad prevalence rates utelised for Inference 3 were derived from Ward et al. (2012); Refshauge and Kalisch (2012) and is summarised in Table S9. Again the reader is cautioned that these are very broad estimates of prevalence rates and are generalised over many factors including lifestyle, genetic and demographic. These prevalence rates also do not take into account participants who develop other forms of dementia or any other neuropsychological disorders. The authors acknowledge there are several factors which the models presented in the manuscript do not account for. As the BLME models and inference derivation presented in this paper are the first of its kind, the objective of Inference 3 is to demonstrate probable diagnosis trajectories conditional on very broad, non-group specific prevalence rates. In order to account for gender and APOE $\varepsilon 4$ status and develop diagnosis trajectories specific to these groups, prevalence rates across ages 65-85 specific to these groups is required, which unfortunately is difficult to find in literature. Figure S10 are the disease trajectories for models (3) in the manuscript applied on male, female, APOE $\varepsilon 4$ carriers and non carriers groups separately, for the ventricle and hippocampus models. We assumed the same prevalence rates as in the manuscript.

12

BMJ Open: first published as 10.1136/bmjopen-2016-012174 on 7 February 2017. Downloaded from http://bmjopen.bmj.com/ on June 13, 2025 at Agence Bibliographique de Enseignement Superieur (ABES)

Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.	Enseignement Superieur (ABES) .	BMJ Open: first published as 10.1136/bmjopen-2016-012174 on 7 February 2017. Downloaded from http://bmjopen.bmj.com/ on June 13, 2025 at Agence Bibli
--	---------------------------------	---

ographique de

Age	HC	MCI	AD
60	0.945	0.037	0.018
65	0.917	0.055	0.028
70	0.859	0.096	0.045
75	0.592	0.333	0.075
80	0.518	0.357	0.125
85	0.466	0.301	0.203

Table S9: Broad prevalence rates for healthy control (HC), mild cognitive impaired (MCI) and Alzheimer's disease taken from Ward et al. (2012); Refshauge and Kalisch (2012). These rates do not account for any lifestyle, demographic and genetic factors as well as other forms of dementia and neuropsychological disorders which are known to affect prevalence rates.

References

- D. Bates, M. Mächler, B. Bolker, and S. Walker. Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*, 2014a.
- D. Bates, M. Maechler, B. Bolker, and S. Walker. *Ime4: Linear mixed-effects models using Eigen and S4*, 2014b. URL http://CRAN.R-project.org/package=lme4. R package version 1.1-7.
- J. Bernal-Rusiel, D. N. Greve, M. Reuter, B. Fischl, and M. R. Sabuncu. Statistical analysis of longitudinal neuroimage data with Linear Mixed Effects models. *NeuroImage*, 66:249–60, February 2013. doi: 10.1016/j.neuroimage.2012.10.065.
- R. R. Corbeil and S. R. Searle. Restricted maximum likelihood (REML) estimation of variance components in the mixed model. *Technometrics*, 18(1):31–38, 1976.
- D. J. MacKay. Information theory, inference and learning algorithms. Cambridge University Press, 2003.
- J. C. Pinheiro. Topics in Mixed Effects Models. PhD thesis, UNIVERSITY OF WISCONSIN–MADISON, 1994.
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013. URL http://www.R-project.org/.
- A. RefshaugeandD. Kalisch.DementiainAustralia,2012.URLhttp://www.aihw.gov.au/WorkArea/DownloadAsset.aspx?id=10737422943.Cat.no. AGE70.
- H. Rue, S. Martino, and N. Chopin. Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the royal statistical society: Series b (statistical methodology)*, 71(2): 319–392, 2009.
- N. H. Timm. 2 multivariate analysis of variance of repeated measurements. Handbook of statistics, 1:41-87, 1980.
- W. Wang and A. Gelman. Difficulty of selecting among multilevel models using predictive accuracy. *Statistics at its Interface*, 7(1):1–88, 2014.
- A. Ward, H. M. Arrighi, S. Michels, and J. M. Cedarbaum. Mild cognitive impairment: disparity of incidence and prevalence estimates. *Alzheimer's & Dementia*, 8(1):14–21, 2012.

BMJ Open: first published as 10.1136/bmjopen-2016-012174 on 7 February 2017. Downloaded from http://bmjopen.bmj.com/ on June 13, 2025 at Agence Bibliographique de I Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.



Figure S10: Male, female, APOE ε 4 carriers and non-carriers diagnosis trajectories for ventricle (top) and hippocampus (bottom) model. Volume quantiles X1, X2, X3 and X4 denote 75-100th, 50-75th, 25-50th and 15-25th quantiles respectively.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml