PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Comparisons of neurodegeneration over time between healthy ageing and Alzheimer's disease cohorts via Bayesian inference
AUTHORS	Cespedes, Marcela; Fripp, Jurgen; McGree, James; Drovandi, Christopher; Mengersen, Kerrie; Doecke, James

VERSION 1 - REVIEW

REVIEWER	Joseph C. Cappelleri Pfizer Inc, United States I am an employee and stockholder of Pfizer Inc.
REVIEW RETURNED	25-Apr-2016

GENERAL COMMENTS	Despite fulfilling most of the checklist items (the two not met should easily be filled upon revision), and the high technical exposition of the manuscript, this technically advanced paper would be a better fit for a journal that regularly publishes on applications of statistics in medicine than it would for the general readership of BMJ Open. Suitable alterative journals include (among others) Statistics in Medicine and Statistical Methods in Medical Research, as well as methodologically-based journal in neuroscience.

REVIEWER	Bin Cheng Columbia University, USA
REVIEW RETURNED	17-May-2016

GENERAL COMMENTS	Major comments: 1. Since the outcome has value between 0 and 1, it is not a good idea to model it using linear mixed model. 2. The authors did not spent effort to investigate nonlinear time trend. The model assumed a linear model. 3. The authors did not convinced me why Bayesian version is necessary. It seems to me that linear mixed effects model can do all they claimed. 4. Page 5, line 53, I_r should be I_n, where n is the total sample size. 5. Page 7, line 12, should be $\beta_4 < \beta_5$, not $\beta_3 < \beta_5$.

REVIEWER	Michael Hornberger Norwich Medical School, Norwich, UK
REVIEW RETURNED	17-Jun-2016

GENERAL COMMENTS	The current study by Cespedes and colleagues explored a Bayesian

approach towards longitudinal imaging analysis of healthy controls, mild cognitive impairment (MCI) and Alzheimer's disease (AD) patients. The authors investigates the hippocampus and ventricles volumes for these 3 groups of patients retrospectively on a large longitudinal dataset. The authors find that Bayesian linear mixed effect models has its values in determining overall diagnostic group trajectories over time, prediction of conversion on a single case basis and how diagnosis trajectories interact with age effects. Overall, this an excellent article exploring how Bayesian frameworks can contribute to better prediction of Alzheimer conversion on a group and single case level. The authors make very clear the advantages and disadvantages of the new technique, however it emerges from the study that Bayesian techniques have great advantages over conventional used linear mixed effect models. I have only a few comments for the authors which will hopefully improve the manuscript further:

• It would have been very useful if the authors also would have done a conventional mixed linear effects analysis. This would have allowed comparing Bayesian and non-Bayesian mixed linear approaches directly. Such an approach would be of great clinical utility to determine the additional benefits of the Bayesian approach.

• The authors limit their analysis to the hippocampus and the ventricle, which is fine. However, such region-of-interest (ROI) approaches can sometimes miss the best classification areas. The authors should therefore consider mentioning that their ROI approach has certain limitations. Alternatively, if possible, they could include other brain areas, in particular to demonstrate that the longitudinal changes are not globalised across the whole brain.

• Some of the results are presented lateralised (eg. left ventricle, right hippocampus) while others seem to be presented as bilateral. It would be helpful to have more consistency across the results by having first the bilateral and then the lateralised findings in the figures and tables.

• Please check/clarify the following sentence in the Methods: "three individuals progressed directly from HC to AD, eight people progressed from HC to AD"

REVIEWER	Juan M. Górriz University of Granada Spain
REVIEW RETURNED	21-Sep-2016

GENERAL COMMENTS	The paper is very interesting although the use of Bayesian Inference based CAD does not preserve the system from overfitting. The number of variables is quite huge. In addition, I recommend to discuss the outcome of the system on the ADNI database, the results on this DB will be the same as the ones achieved in this study? I also recommend the use of some validation strategies such as K-Fold, LOO, etc. to preserve the system from overfitting. Some references could be added as well: L. Khedher, et al.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) Andrés Ortiz, et al. Exploratory graphical models of functional and structural connectivity patterns for Alzheimer's Disease diagnosis.

	Front. Comput. Neurosci. 2015 (2015)
	Juan Eloy Arco, et al. Improving short-term prediction from MCI to AD by applying searchlight analysis. ISBI 2016: 10-13
	Z. Dai, et. al. "Discriminative analysis of early Alzheimer's disease using multi-modal imaging and multi-level characterization with multi-classifier (M3).," NeuroImage, vol. 59, no. 3, pp. 2187 – 2195, 2012
	J.E. Arco, et al. "Short-term Prediction of MCI to AD conversion based on Longitudinal MRI analysis and neuropsychological tests.," in Innovation in Medicine Healthcare 2015, 2015, pp. 385–394.
	to complete the introduction and the recently reported state-of the art. In these references the machine learning based approach was applied to MCI prediction on the ADNI dataset and overfitting was relieved by the use of cross-validation. Finally, authors should discuss the different pattern evolution obtained from non-converter MCIs vs converters subjects (if this information is available) and how the volume changes are differently affected on these classes.

VERSION 1 – AUTHOR RESPONSE

Reviewer (1) Joseph C. Cappelleri.
Despite fulfilling most of the checklist items (the two not met should easily be filled upon revision), and the high technical exposition of the manuscript, this technically advanced paper would be a better fit for a journal that regularly publishes on applications of statistics in medicine than it would for the general readership of BMJ Open.

Suitable alternative journals include (among others) Statistics in Medicine and Statistical Methods in Medical Research, as well as methodologically-based journal in neuroscience.

Response: We thank the first reviewer Joseph C. Cappelleri for his comment.
In this paper, we present the results and provide new insights into clinically relevant research questions facilitated by the application of advanced statistical methodologies (that is, Bayesian linear mixed effects modelling) to analyse longitudinal neuroimaging data. Thus, our target audience is practitioners, and it seems that BMJ is an appropriate journal. The alternative journals suggested by the referee may be more appropriate if we were targeting statisticians where the focus would be much more on the methodological developments in statistics.
###
Reviewer (2) Bin Cheng
Major comments:
1. Since the outcome has value between 0 and 1, it is not a good idea to model it using linear mixed model.
Response: We thank the second reviewer Bin Cheng for his comments and corrections.

In general the referee is correct; typically one wouldn't consider a linear mixed model for constrained data. However, in this study, despite the data being constrained in (0, 1), no data points are located near either boundary. Indeed provided the statistical modelling assumptions are valid (see supplementary material), one can model such data via a linear mixed model. Of course, one should carefully interpret the results and be aware of the limitations of the model. A discussion about this has been included in the method Section 3, page 4 line 54 as;

"While our response values are constrained to 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."

2. The authors did not spent effort to investigate nonlinear time trend. The model assumed a linear model.
Response: Non-linear time trends, such as Age^2, Age^3 and Age^4 have now been investigated and models were compared by the posterior model probability, please see the supplementary material. The model with the highest posterior model probability for both ventricle and hippocampus models consists of Age as a linear terms, and this justifies our choice for a linear model. Discussion on exploration of non-linear time trends is included in Section 3.2 on page 5 line 35, as

"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."

3. The authors did not convinced me why Bayesian version is necessary. It seems to me that linear mixed effects model can do all they claimed.
Response: A classical LME of the same form as expression (3) was performed (sentence added to Section 3.2 on page 6 line 11), and the same three research questions presented in the paper were attempted, please refer to Section 4 in the Supplementary material for full results.

We find that A) our credible intervals support the results from the hypothesis test, however the classical approach does not allow for probability statements of group ordering (see Section 4.1 on page 9 line 10 of the revised paper).

B) There are strong similarities between the Bayesian and classical LME of participant rankings, including the eight HC to MCI converters on both regions. However under the classical LME, the distribution of ranks for participants cannot be derived, this discussion is further expanded in Section 4.2 on page 9 line 27.

And C) region specific diagnosis trajectories on volume quantiles cannot be performed with a classical LME model, as was done with the BLME, Section 3.5 on page 8 line 15 explains this in further detail. Further discussion to summarise these points in comparison to the BLME presented in the paper is provided in the amended manuscript in Section 5 on page 12.

4. Page 5, line 53, I_r should be I_n, where n is the total sample size.
Response: Corrected I_r to I_n where n is the total sample size.

5. Page 7, line 12, should be $\beta_4 < \beta_5$, not $\beta_3 < \beta_5$.
Response: In expression (4) and sentence below corrected to $\beta_4^m - \beta_5^m < 0$.
###
Reviewer (3) Michael Hornberger
The current study by Cespedes and colleagues explored a Bayesian approach towards longitudinal imaging analysis of healthy controls, mild cognitive impairment (MCI) and Alzheimer's disease (AD) patients. The authors investigates the hippocampus and ventricles volumes for these 3 groups of patients retrospectively on a large longitudinal dataset. The authors find that Bayesian linear mixed effect models has its values in determining overall diagnostic group trajectories over time, prediction of conversion on a single case basis and how diagnosis trajectories interact with age effects.

Overall, this an excellent article exploring how Bayesian frameworks can contribute to better prediction of Alzheimer conversion on a group and single case level. The authors make very clear the advantages and disadvantages of the new technique, however it emerges from the study that Bayesian techniques have great advantages over conventional used linear mixed effect models. I have only a few comments for the authors which will hopefully improve the manuscript further:

Response: We thank the 3rd reviewer Michael Hornberger, for his positive and helpful comments about this paper.

• It would have been very useful if the authors also would have done a conventional mixed linear effects analysis. This would have allowed comparing Bayesian and non-Bayesian mixed linear approaches directly. Such an approach would be of great clinical utility to determine the additional benefits of the Bayesian approach.
Response: A classical LME of the same form as expression (3) was performed (sentence added to Section 3.2 on page 6 line 11), and the same three research questions presented in the paper were attempted, please refer to Section 4 in the Supplementary material for full results.

We find that A) our credible intervals support the results from the hypothesis test, however the classical approach does not allow for probability statements of group ordering (see Section 4.1 on page 9 line 10 of the revised paper).

B) There are strong similarities between the Bayesian and classical LME of participant rankings, including the eight HC to MCI converters on both regions. However under the classical LME, the distribution of ranks for participants cannot be derived, this discussion is further expanded in Section 4.2 on page 9 line 27.

And C) region specific diagnosis trajectories on volume quantiles cannot be performed with a classical LME model, as was done with the BLME, Section 3.5 on page 8 line 15 explains this in further detail. Further discussion to summarise these points in comparison to the BLME presented in the paper is provided in the amended manuscript in Section 5 on page 12.

• The authors limit their analysis to the hippocampus and the ventricle, which is fine. However, such region-of-interest (ROI) approaches can sometimes miss the best classification areas. The authors should therefore consider mentioning that their ROI approach has certain limitations. Alternatively, if possible, they could include other brain areas, in particular to demonstrate that the longitudinal changes are not globalised across the whole brain.
Response: We agree, our analysis which focuses on specific ROI's (lateral ventricles and hippocampus) limits the inference specific to those areas, irrespective of the entire complex neural/region interactions. This limitation is re-iterated in Section 2 as follows

"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."

Also on page 13 in Section 5 line 10 suggests future work or expansion of application and methodology to account for multiple brain regions simultaneously over time. However, as the main aspect of our work is region and disease specific, longitudinal changes analysed in Bayesian linear mixed effects model, inclusion of many regions (more than two), or additional spatial effects is beyond the scope of our paper.

• Some of the results are presented lateralised (eg. left ventricle, right hippocampus) while others

seem to be presented as bilateral. It would be helpful to have more consistency across the results by having first the bilateral and then the lateralised findings in the figures and tables.

Response: Results Section 4 has been revised, for consistency across results as suggested.

• Please check/clarify the following sentence in the Methods: "three individuals progressed directly from HC to AD, eight people progressed from HC to AD"

Response: This was a typo, sentence in Section 3.4 has been corrected.
###
Reviewer (4) Juan M. Górriz
The paper is very interesting although the use of Bayesian Inference based CAD does not preserve the system from overfitting.

I also recommend the use of some validation strategies such as K-Fold, LOO, etc. to preserve the system from overfitting.
Response: We thank the 4th reviewer, Juan M. Gorriz for his helpful comments.
Two adaptations of leave-one-out cross validation assessments have now been performed on both the ventricle and hippocampus models. Due to the nature of hierarchical model, we assessed out of sample validation for both a set of observations for each individual, as well as within individual observations. Full results are included in the Supplementary materials and discussions related to these results were included in Section 3.2 on page 5 line 49, as shown below
"In addition to the residual 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.54 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 a low MSE values, which supports our model choice, refer to Supplementary material for full details."

The number of variables is quite huge. In addition, I recommend to discuss the outcome of the system on the ADNI database, the results on this DB will be the same as the ones achieved in this study?

In this work we lay the foundation for statistical inference through a BLME model to address research questions in a clinical setting, and this is demonstrated in the application of the AIBL study. We note that while possible extensions and applications to different longitudinal studies such as ADNI were suggested in the manuscript, the focus of our manuscript is to demonstrate an alternative approach to address important clinical questions through a BLME model.
We encourage readers to replicate our analysis either on the simulated data provided with the full R code, or by direct application to their respective clinical study. Furthermore, as our manuscript (approx. 6, 500 words) is already well beyond the BMJ Open word recommendation limit (4,000 words), further analysis and application to an independent data set, such as ADNI would be better suited as a follow-on manuscript.
Some references could be added as well:

L. Khedher, et al.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).

Andrés Ortiz, et al. Exploratory graphical models of functional and structural connectivity patterns for Alzheimer's Disease diagnosis. Front. Comput. Neurosci. 2015 (2015).

Juan Eloy Arco, et al. Improving short-term prediction from MCI to AD by applying searchlight analysis. ISBI 2016: 10-13 – Searchlight.

Z. Dai, et. al. "Discriminative analysis of early Alzheimer's disease using multi-modal imaging and multi-level characterization with multi-classifier (M3).," NeuroImage, vol. 59, no. 3, pp. 2187 – 2195, 2012

J.E. Arco, et al. "Short-term Prediction of MCI to AD conversion based on Longitudinal MRI analysis and neuropsychological tests.," in Innovation in Medicine Healthcare 2015, 2015, pp. 385–394.

to complete the introduction and the recently reported state-of the art. In these references the machine learning based approach was applied to MCI prediction on the ADNI dataset and overfitting was relieved by the use of cross-validation.

We thank the referee for the references, the five suggested references were included in Section 1, the introduction of the manuscript as shown below, to provide motivation on recently reported state of the art analysis on HC, MCI and AD groups.
"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"
Finally, authors should discuss the different pattern evolution obtained from non-converter MCIs vs converters subjects (if this information is available) and how the volume changes are differently affected on these classes.

Sections 3.4 and 4.2 have been revised to describe all converters in our analysis.
The volumetric rate of change for the ventricle and hippocampus regions on all participants (including MCI and converters) is presented in Figure 3 which illustrates the patterns of converters and non-converters over all diagnosis groups for the ventricle and hippocampus regions. However as the box plots show a large level of uncertainty in the estimated rate of volume change for each participants given our unbalanced data, our analysis also focused further on two MCI to AD converters which were at the 15th quantile extreme in both regions, and showed increasing probability or remaining in the worst 15th percentile.

This is discussed in detail in Section 4.2 on page 10 starting on line 20, in particular
"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 15th 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 15th quantile are 0.80 and 0.66 for the ventricle and 0.54 and 0.69 for the hippocampus regions."

VERSION 2 – REVIEW

REVIEWER	Michael Hornberger Professor of Dementia Research Department of Medicine, Norwich Medical School University of East Anglia, UK
REVIEW RETURNED	27-Oct-2016

GENERAL COMMENTS	The authors have addressed all my comments sufficiently.

REVIEWER	Juan M. Górriz University of Granada, Spain
REVIEW RETURNED	03-Nov-2016

GENERAL COMMENTS	Authors have addressed all the questions raised by this reviewer