PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	The Association of Brain Amyloidosis with the incidence and frequency of Neuropsychiatric Symptoms in ADNI - a multisite observational cohort study
AUTHORS	Goukasian, Naira; Hwang, Kristy; Romero, Tamineh; Grotts, Jonathan; Do, Triet; Groh, Jenna; Bateman, Daniel; Apostolova, Liana

VERSION 1 – REVIEW

REVIEWER	Yun Zhou Huazhong University of Science and Technology, Wuhan, China
REVIEW RETURNED	01-Jul-2019

GENERAL COMMENTS	-Methods Section, Subjects Subsection. The subjects included in this study were from ADNI-1, ADNI-GO, and ADNI-2. The author described that “ADNI expanded enrollment criteria with the launch of ADNI-GO in 2009 and enrolled 200 additional subjects with early amnestic MCI (EMCI)” and “ADNI-2 added approximately 650 newly enrolled subjects”. I am wondering that whether there were parts of the subjects receiving more than once examination or not? If yes, it should be considered when analyzing. -Discussion Section, paragraph 3 and 4. The authors should interpret the findings they observed rather than just repeating the results here.

REVIEWER	Paul Rosenberg Johns Hopkins University School of Medicine
REVIEW RETURNED	17-Jul-2019

GENERAL COMMENTS	The authors have addressed the issues raised by the prior reviews largely satisfactorily. There remain minor issues including: 1. The somewhat paradoxical finding that amyloid positivity in demented ADNI participants was associated with lower prevalence of apathy, agitation, and appetite changes. The authors explanation (response to reviewer 1 item 3) is not entirely satisfactory in my opinion. Instead, take note of the characteristics of the group with more symptoms: demented without amyloid. It is likely that most of these participants are suffering from a non-AD dementia. If this includes DLB (which usually yields equivocal or negative amyloid PET scans) and FTD (negative amyloid PET scans), that might explain these results. 2. Discussion: ungrammatical sentence “Our Cox proportional hazard regression

	model further demonstrated that in the MCI due to AD develop apathy, anxiety, and agitation - three of the earliest and most pervasive NPS in the MCI stage" probably meant to say "...participants with MCI due to AD develop apathy [etc.] earlier...." 3. I would delete Figure 1 and just make mention of the null findings about frequency in the text. Usually one doesn't even bother with presenting NPI frequencies by themselves but in this context – mix of NPI and NPI-Q – worth mentioning.

REVIEWER	Charles Malpas The University of Melbourne, Melbourne, Australia
REVIEW RETURNED	04-Aug-2019

GENERAL COMMENTS	In this large ADNI paper, the authors investigate the relationship between amyloid pathology and neuropsychiatric symptoms. This is an interesting paper and contributes to the understanding of the broader phenotype of Alzheimer's disease. Overall, I feel this paper is appropriate for publication, conditional on the resolution of some points below. As such, my review will focus on new issues that I feel would be worthwhile addressing. It appears that the paper has already been through 1 phase of review. The introduction and discussion sections are generally clear and place the study in context. It would be good if the authors could explicitly state the difference between the NPI and NPI-Q given that both were used in the study. This will help the reader understand the strengths and limitations of each. The authors have corrected the significance (p) values using a false discovery rate method. The specific method should be mentioned (there are several approaches to FDR). It is also not clear over what set of p values the FDR was derived. I assume this was all p values in the paper (the most appropriate choice) but this should be made explicit. Was q set at 5%? If so, this should be specified in the methods. Are the p values reported in the results section 'FDR corrected'? The authors refer to statistical findings that were above the critical alpha level of 5% as 'trending'. This language is only acceptable if values marginally below the critical alpha are considered 'trending away from significance'. If the authors are using the Neyman-Pearson framework (which their use of a critical alpha and FDR would imply) then the value of the p value is only of value if it is below or above the critical alpha. No meaning can be derived from the magnitude of the p value, as the Neyman-Pearson framework specifically ignores the strength of evidence against the null. Findings that were above critical alpha should just be referred to as not statistically significant. The cox regression models were referred to as both 'corrected' and 'controlling' for other variables. Please change this to 'adjusting for' to avoid confusion. Only experimental designs can meaningfully control for confounders. The authors should report the assumption checks for the Cox PH models.

	The authors have used backwards logistic regression in their analysis. While backwards selection is the most appropriate of all statistical variable selection methods, it is important to acknowledge that performing variable selection and inference on the same dataset is problematic and runs the risk of poor bias-variance trade-off. The authors should briefly mention this in the discussion and acknowledge the need for external confirmation of these findings.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

1. Methods Section, Subjects Subsection. The subjects included in this study were from ADNI-1, ADNI-GO, and ADNI-2. The author described that "ADNI expanded enrollment criteria with the launch of ADNI-GO in 2009 and enrolled 200 additional subjects with early amnestic MCI (EMCI)" and "ADNI-2 added approximately 650 newly enrolled subjects". I am wondering that whether there were parts of the subjects receiving more than once examination or not? If yes, it should be considered when analyzing.

Those enrolled in ADNI-1 only had the NPI-q available, while those enrolled in ADNI-2 and ADNI-GO were administered the full version. There was no overlap. Each patient had either NPI or NPI-q for each visit that was analyzed. This is further explained under Methods: Neuropsychiatric Data as follows:

Those enrolled in ADNI-1 only had the NPI-Q available, while those enrolled in ADNI-2 and ADNI-GO were administered the full version. NPI or NPI-Q data from the baseline and all annual visits were obtained from LONI IDA on November 3, 2015. Each patient had either NPI or NPI-Q for each visit that was analyzed.

2. Discussion Section, paragraph 3 and 4. The authors should interpret the findings they observed rather than just repeating the results here.

We have an interpretation following the short summary of the results which we feel is necessary due to the very dense results section. The interpretation reads as follows:

"Taken together our data seem to indicate that both prevalent agitation, and apathy, and incident agitation, irritability and apathy are predictive of faster functional decline and loss of independence among amyloid positive and negative MCI."

Reviewer: 2

1. The somewhat paradoxical finding that amyloid positivity in demented ADNI participants was associated with lower prevalence of apathy, agitation, and appetite changes. The authors explanation (response to reviewer 1 item 3) is not entirely satisfactory in my opinion. Instead, take note of the characteristics of the group with more symptoms: demented without amyloid. It is likely that most of these participants are suffering from a non-AD dementia. If this includes DLB (which usually yields equivocal or negative amyloid PET scans) and FTD (negative amyloid PET scans), that might explain these results.

Thank you for the suggestion. The following has been added to our discussion section:

“Contrary to our expectations we found that amyloid negative dementia subjects have higher frequency of apathy, agitation and appetite changes compared to the amyloid positive dementia group. One possible explanation is that previous studies on the prevalence of NPS in AD dementia have not included biomarker validation. This means that AD phenocopies with amnestic presentation were included as AD cases. ADNI is the first large scale observational study that included amyloid PET as a biomarker. Thus, for the first time we have the opportunity to investigate the frequency of NPS in biomarker validated AD dementia and compare that to AD phenocopies. What we find here suggests that these previous reports might have overestimated the true prevalence of some NPS in AD dementia due to including amyloid negative AD phenocopies in the AD dementia group. An alternative explanation is that ADNI subjects who are amyloid negative have other neurodegenerative disorders such as frontotemporal dementia, dementia with Lewy bodies, argyrophilic grain disease, hippocampal sclerosis, etc. These other pathologic entities could also explain the increased frequencies of NPS.”

2. Discussion: ungrammatical sentence “Our Cox proportional hazard regression model further demonstrated that in the MCI due to AD develop apathy, anxiety, and agitation - three of the earliest and most pervasive NPS in the MCI stage” probably meant to say "...participants with MCI due to AD develop apathy [etc.] earlier...."

Thank you for the suggestion. The sentence now reads:

“Our Cox proportional hazard regression model further demonstrated that participants with MCI due to AD develop apathy, anxiety, and agitation - three of the earliest and most pervasive NPS in the MCI stage, earlier compered to amyloid negative MCI”

3. I would delete Figure 1 and just make mention of the null findings about frequency in the text. Usually one doesn’t even bother with presenting NPI frequencies by themselves but in this context – mix of NPI and NPI-Q – worth mentioning.

Thank you for the suggestion.

Figure 1 has been removed and the rest of the figures were adjusted accordingly.

Reviewer: 3

1. It would be good if the authors could explicitly state the difference between the NPI and NPI-Q given that both were used in the study. This will help the reader understand the strengths and limitations of each.

The following clarification was added to the Methods section:

“Interviewers ask structured questions about the presence and severity (as well as frequency in the full version of the NPI but not in NPI-Q) of the symptoms in the past month to the study partner. “

2. The authors have corrected the significance (p) values using a false discovery rate method. The specific method should be mentioned (there are several approaches to FDR). It is also not clear over what set of p values the FDR was derived. I assume this was all p values in the paper (the most appropriate choice) but this should be made explicit. Was q set at 5%? If so, this should be specified in the methods. Are the p values reported in the results section ‘FDR corrected’?

The correction method we used was the Benjamini and Hochberg false discovery rate (FDR). The p-values were adjusted independently for each set of analysis. In addition, every model was adjusted for multiple comparison independently.

The following was adjusted in the manuscript:

"Cox regression models were evaluated for proportional hazard assumption. All p-values were adjusted for multiple comparison correction using Benjamini & Hochberg false discovery rate (FDR) correction. All statistical tests were two-sided and a q-value less than 0.05 was considered statistically significant."

3. The authors refer to statistical findings that were above the critical alpha level of 5% as 'trending'. This language is only acceptable if values marginally below the critical alpha are considered 'trending away from significance'. If the authors are using the Neyman-Pearson framework (which their use of a critical alpha and FDR would imply) then the value of the p value is only of value if it is below or above the critical alpha. No meaning can be derived from the magnitude of the p value, as the Neyman-Pearson framework specifically ignores the strength of evidence against the null. Findings that were above critical alpha should just be referred to as not statistically significant.

Thank you for the suggestion. "Trending" data is no longer discussed in the results section.

4. The cox regression models were referred to as both 'corrected' and 'controlling' for other variables. Please change this to 'adjusting for' to avoid confusion. Only experimental designs can meaningfully control for confounders.

All Cox regression models have been changed and are now referred to as 'adjusting' or 'adjusted' for.

5. The authors should report the assumption checks for the Cox PH models.

As suggested, the following was added to the manuscript:

"All Cox regression models were evaluated for proportional hazard assumption and there was no evidence that the models did not meet required assumptions (q-value>0.081)."

6. The authors have used backwards logistic regression in their analysis. While backwards selection is the most appropriate of all statistical variable selection methods, it is important to acknowledge that performing variable selection and inference on the same dataset is problematic and runs the risk of poor bias-variance trade-off. The authors should briefly mention this in the discussion and acknowledge the need for external confirmation of these findings.

Thank you for the suggestion. The following has been added to our discussion:

"Since the model selection and model building were done in the same dataset, our results are considered explanatory. Further studies are warranted to confirm our results".

VERSION 2 – REVIEW

REVIEWER	Paul B. Rosenberg, M.D. Johns Hopkins University School of Medicine USA
REVIEW RETURNED	24-Oct-2019

GENERAL COMMENTS	The authors have addressed the issues raised in my prior review to my satisfaction.

REVIEWER	Charles Malpas Senior Research Fellow, Clinical Outcomes Research Unit, Department of Medicine, Royal Melbourne Hospital, The University of Melbourne, Australia
REVIEW RETURNED	08-Oct-2019

GENERAL COMMENTS	I thank the authors for responding to my queries. Please address one final typographical error. In the sentence "Since the model selection and model building were done in the same dataset, our results are considered explanatory. Further studies are warranted to confirm our results" please change explanatory to exploratory.