Reviewer 1 Comments for the Author...

In this really interesting study authors are using innovative AI algorithms to estimate chronological age of different group subjects from their neuroimaging data. Interestingly, they are comparing MRI-BAG and FDG PET-BAG as imaging biomarkers of neurodegeneration. Moreover, they explore the associations of these two parameters with cognitive performance and Alzheimer pathology (amyloid and tau), and an estimation of the probability of cognitive outcome value by means of these parameters.

Unfortunately, the study contains major **methodological flaws related to the statistical analysis. Authors extracted an equally sized subsample of stable subjects matched by age and sex to the cohort of all decliners (both in CN+SCDADNI and MCIADNI subjects)**. Instead of conventional logistic regression, this approach needs a **conditional logistic regression analysis due to the presence of matching**. Moreover, a *case control study* like it is proposed here (similar number of subjects matched by age and sex) allows to estimate the Odds Ratio but not the probability of cognitive outcome. For this purpose, a **study of cohorts (using the whole population available) would allow to estimate the real probability of conversion**. Therefore, it should be discussed with their statistician if this study have to include the whole population **adjusted by age and sex as covariates and use a conventional logistic regression analysis** to model the relationship between BAG from the significant imaging modality and cognitive outcome.

Additional comments:

Some of the statistical methods are described in the specific sections of comparisons and correlations performed. However, there is a lack of **section where all the statistical analysis should be detailed**. Moreover, the **description of r of Pearson and rho of Spearman analysis only appears at the results section**.

To derive a BAG cutoff, authors intercept of the logistic regression curve at 50% probability. However, the use of this cutoff might be arbitrary. Instead, you should use **AUC ROC graphs for optimal cutoff of maximal sensitivity and specificity**.

**Please present the mean difference and confidence interval instead of t statistic in the respective paragraphs of the result section.**

**PPV and NPV can´t be estimated in case control studies unless you used specific**

**methods (i.e. Mercaldo).**

Please include **confidence intervals for all indexes described**. **???**

The term marginal is frequently misleading in the text. It should be used **“very weak but almost significant”**. Take into account that very big sample sizes lead to a lot of irrelevant but statistically not significant findings.

Use **“FDG PET BAG” instead of “FDG PET”** when referring specifically to this particular parameter, (FDG PET is a more general term which offers some additional important parameters). Example: “… both MRI- and FDG-PET-BAG significantly predicted cognitive outcome in MCIADNI, while only FDG-PET predicted cognitive outcome in CN+SCDADNI (see Tables SM2-SM5 for estimates of logistic regression in sample 1 analyses using only one imaging modality).”

Please use **“and” or “or“ instead of a “slash”** in the text. These are two

examples:

• “… cognitively normal (CN)/had subjective cognitive decline (SCD)”. You can use CN/SCD as an acronymous but after an explanatory text without symbols.

• “… with cognitive performance/AD neuropathology in these cohorts.”

Please explain why you didn’t **consider to add the Ratio CSF AB-pTau to the single AB or pTau in CSF**.

Some discussion about the specific **cortical areas that have been shown to be**

**prognostic in other studies (ie. Posterior Cingulate, Parietal for FDG-PET, and**

**hippocampus for MRI) are missed**. Besides, a comparison with other papers which

explored the brain **hypometabolism in FDG PET and cortical atrophy in MRI should**

**be convenient.**

Reviewer 2 Comments for the Author...

Doering et al. applied machine learning algorithms to structural MRI and FDG-PET images of n=376 elderly subjects with subjective cognitive decline (SCD) and without to calculate brain age gap (BAG). The authors associated MRI- and PET-derived BAG with some cognitive tests, biomarkers of Alzheimer’s disease (AD), and with a cognitive outcome in 2 years. They report correlations between BAG and amyloid-beta in cerebrospinal fluid in the above subjects and in patients with mild cognitive impairment (MCI), as well as between BAG and cognitive performance in MCI patients. Furthermore, PET-derived BAG predicted cognitive deterioration in SCD+healthy subjects, while MRI-derived BAG predicted cognitive deterioration (to dementia) in MCI patients. Doering et al. conclude that BAG can be reliably estimated from FDG-PET and MRI images. Whereas PET-derived BAG is more sensitive to cognitive deterioration in subjects without objective cognitive impairment, MRI-derived BAG is indicative of impending dementia in patients with MCI.

Major criticism

1) Application of the results. As the authors themselves note, BAG has been defined so far using MRI data. This is understandable, since **FDG-PET is not indicated in cognitively healthy subjects**. Hence, application of **PET-derived BAG is basically limited to clinical trials (and academic studies).** Specifically, the authors note that “FDG-PET BAG could complement the identification of at risk individuals, as individuals with a BAG below our proposed cuttoff are unlikely to develop cognitive impairment within two years”. This statement is based on findings in an external cohort of SCD subjects (“DELCODE”), where sens, spec, PPV and NPV of 88%, 34%, 13%, and 96% are reported. Yet, this cohort include n=80 cognitively stable and only n=8 cognitively deteriorated subjects. Thus, there is **per definition a bias toward high NPV**. Remarkably, in an equally weighted (n=30 stable vs. n=30 decliners) subset of the initial cohort, i.e., the cohort that was used to derive PET-BAG, sens, spec, PPV, and NPV were only 70%, 67%, 68%, and 69%. I’m afraid, **these results are not sufficient to recommend PET-derived BAG for the use in clinical trials**. Consistently with this **limited predictive power of PET-derived BAG, it either did not correlate or correlated only marginally (r=-0.100, p=0.06) with the cognitive tests.**

2) Irrespective of the above, what is the point of using **PET-derived BAG instead of PET itself as marker of cognitive decline**? See e.g., Scheef et al., 2012. The former requires rather complex analyses including machine learning, while analytical pipelines of FDG-PET data are well established. In the same vein, **MRI features alone might predict cognitive decline in SCD subjects, see Ebenau et al., 2022**. To summarize the first two comments, I question the utility of BAG in general and PET-derived BAG in particular as marker of cognitive decline in clinical trials. To justify this application, the authors should **compare BAG with established regional features of FDG-PET, MRI images (e.g., hippocampal volume), and with chronological age as reference**. This is equally true for MCI patients. Otherwise, **another application of (in particular PET-derived) BAG should be proposed.**

3) The main study cohort represents a mixture of SCD subjects and cognitively healthy subjects without cognitive complaints. As the authors themselves note, SCD subjects are more likely to develop MCI and dementia due to AD, for a recent meta-analysis see Pike et al., 2022. Thus, BAG is biased towards disease-related acceleration. The **authors should treat these groups separately or exclude subjects with SCD**. Inclusion criteria should be clearly stated. In particular, what is the **status of cognitively healthy subjects regarding the AD biomarkers?**

Further comments

1) It is not plausible that MRI- rather than PET-derived BAG predicted cognitive decline in MCI patients. Numerous studies reported FDG-PET to predict cognitive decline in a more sensitive manner than MRI. The information on chronological age is the same in both BAGs. **How do the authors explain this finding?**

2) **Why only 60+ subjects were included? Accelerated aging to be captured by BAG should begin earlier.**

3) To disentangle effects of potential atrophy on FDG-PET data and to enable a comparison with the literature (Lee et al., 2022), the **PET data should be corrected for partial volume effects (PVE).** Of note, the parcellation into 216 regions should results in a number of regions with a very small volume = significant PVE. So, the results both with and without PVE correction should be presented.

4) How **established and robust is the pipeline that the authors used to calculate BAG?** It is user **friendly and publicly or commercially available** (keeping clinical trials in mind)? How do the results change if e.g., **3- or 6-fold (instead of 5-fold) cross-validation is applied**?

5) **Why the demographic variables, in particular age, are corrected for by default?** BAG is per definition the product of age. Results of the cognitive tests are typically adjusted for the demographic variables (z scores). Superfluous adjustment for variables may lead to spurious associations.

6) Validation in the OASIS sample is described in Methods, **but is absent in Results.**

References

- Ebenau JL, Pelkmans W, Verberk IMW, Verfaillie SCJ, van den Bosch KA, van Leeuwenstijn M, Collij LE, Scheltens P, Prins ND, Barkhof F, van Berckel BNM, Teunissen CE, van der Flier WM. Association of CSF, Plasma, and Imaging Markers of Neurodegeneration With Clinical Progression in People With Subjective Cognitive Decline. Neurology. 2022 Mar 29;98(13):e1315-e1326.

- Pike KE, Cavuoto MG, Li L, Wright BJ, Kinsella GJ. Subjective Cognitive Decline: Level of Risk for Future Dementia and Mild Cognitive Impairment, a Meta-Analysis of Longitudinal Studies. Neuropsychol Rev. 2022 Dec;32(4):703-735.

- Scheef L, Spottke A, Daerr M, Joe A, Striepens N, Kölsch H, Popp J, Daamen M, Gorris D, Heneka MT, Boecker H, Biersack HJ, Maier W, Schild HH, Wagner M, Jessen F. Glucose metabolism, gray matter structure, and memory decline in subjective memory impairment. Neurology. 2012 Sep 25;79(13):1332-9.