Review of: "Coupling between Human Brain Cortical Thickness and Glucose Metabolism from Regional to Connective level: a PET/MRI study"

Auriel Willette[1]

1 University of Iowa

Potential competing interests: No potential competing interests to declare.

Glucose use in the brain and the structural composition of it tend to go down across the lifespan, particularly in synaptically dense areas (e.g., precuneus) or ones that myelinate last in the early part of the lifespan (e.g., prefrontal cortex). A great deal of work has focused on so-called functional connectivity (i.e., autocorrelation between two different areas/networks using FDG-PET, MR-based ASL, etc.) and structural connectivity (i.e., autocorrelation between two different areas/networks using T1-weighted MR imaging). Here, the authors took FDG-PET and MR scans among 138 people, divided them based on age, ran Spearman correlations, and again ran correlations based on age. For positive comments, this report seeks to look at age-related differences in the intersection between structural and functional connectivity. This computational biology approach is difficult to do and can be quite time-consuming. The authors also use an approach that has no a priori assumptions, instead using FDR correction to examine cross-correlations as broadly as possible. Unfortunately, there are major flaws that limit enthusiasm. Perhaps owing to word limit or other considerations, there are some key details about image pre- and postprocessing that are not clear and may lead to bias. This is exacerbated by having inexact age cutoffs, where one group is not entirely middle-aged and the other group is not entirely aged. The results section is unfortunately marred by these decisions and differences in statistical connectivity that do not make sense in the context of aging. For example, functional and structural connectivity are much stronger over the whole brain in the middle-aged + aged group, both in areas where aging does not affect either measure much (e.g., post-central gyrus), but also in areas where biologically there are major differences between the two imaging modalities (e.g., prefrontal cortex). Some portion of the discussion offers useful insight about the results, but based on study limitations the limitations section should be expanded. Further, some claims are not corroborated by what is shown in the results section.

First, on a practical level, this type of work may be comparing apples and oranges. Using 3T MR units, modern T1-weighted sequences achieve ~0.8-1.1mm native resolution in the plane of acquisition. By contrast, PET or PET/CT scanners typically achieve 3-4mm native resolution. While it is possible to downgrade T1 resolution to match that of the PET images, this may not have been done. To be sure, PET and T1-weighted images were coregistered and normalized to a standardized atlas (MNI), but it's unclear whether T1 images were downsampled, PET images were upsampled, or how the different resolution of the two image types were harmonized.

For the introduction, broadly, I do not agree with most of it. To begin, it is oddly grandiose to claim that disruption between

structural and functional connectivity can implicate various disorders like Alzheimer's disease (AD), Traumatic Brain Injury (TBI), and the like. While connectivity pattern differences have been seen for these and other clinical or neurodegenerative conditions, they are not sufficient and do not approach the level of sufficiency needed to validate this claim. Such differences can relate to biomarkers or suggestive indices, but even then regional brain differences for conventional PET or MRI are not sufficient to explain disease-related changes in brain. In another case, these comparisons are considered important because of alleged disconnects between the two types of scans, such as with Baik et al. (2023) in J. Clin. Neurology. I think the authors may be missing the point. Current scan types in PET or MRI are not robust enough to adequately explain variance accounting for controls versus disease populations. In the case of Alzheimer's disease (AD), for example, cognitively unimpaired (CU) vs. AD comparisons might reach roughly 85-90% with a scan modality (e.g., FreeSurfer output), but this same level of performance is often found using cognitive tests. While the variability of these scans do come from two different modalities, their only usefulness relative to cognitive test comparisons is being roughly 4-5 years "ahead" of impending cognitive deficits. Due to the close nature of the two age groups in this study, it's unclear if the current corpus of results takes advantage of this narrow window of time.
Next, "functional connectivity" is a term used almost exclusively with an MR scan modality called resting state functional magnetic resonance imaging (rsfMRI). This connectivity is estimated in dozens to hundreds of acquisitions of the same image over 7-11 minutes, ensuring robust delineation of brain networks using seed-based analysis, Independent Component Analysis, or other approaches. By contrast, 18F-FDG PET typically can be done to get only 5-6 "reads" of the whole brain over 35-45 minutes and inherently has less spatial resolution than rsfMRI. Next, the current study is derivative and follows one published in 2017 by Di et al. using Alzheimer's Disease Neuroimaging Initiative (ADNI) data. There, using participants who had both fMRI and anatomical MRI scans, correlations between scans had small effect sizes and networks found in one scan type did not correspond with the other. Yet, the authors of the current article go against this suggestion of this same study they cited, to indicate that the types of scans (with one using FDG PET, not rsfMRI) could be compatible. How, then, can the authors use this study to suggest a benefit of looking at FDG PET and T1-weighted imaging, which again have completely different acquisition parameters and native scan resolution? Finally, by definition, changes in FDG and cortical thickness occur in different brain regions both in normal aging and AD & related dementias. While looking at connectivity analyses is an interesting premise for FDG and T1-weighted imaging, a cross-sectional approach combined with concerns about group composition (see below) will lead to uneven correlations that unfortunately just reflect the aging process.

For the methods, I have a similar bevy of concerns. First, an exclusion criterion was "neuropsychiatry disorders." Given that a major emphasis in the introduction was learning more about disease pathophysiology, the introduction needs to be tightened up to indicate that neurodegenerative diseases are of interest rather than broad neuropathology. Next, FreeSurfer is used to extract data on brain volume, cortical thickness, local gyrification index, and sulcal depth. Given that FreeSurfer natively produces ~70 ROIs per image type, this quickly becomes problematic even from an FDR standpoint. Specifically, with several FreeSurfer imaging sub-types that are each composed of ~70 ROIs, FDR or FWE correction will render most of these results as non-significant. Further, given that there are several network properties calculated per ROI that expand the number of comparisons made to several hundred, this problem is further aggravated by the study design.

A lack of a priori delineation of hypotheses is fine, but I fear the net may be cast too wide and several interesting results lost to type 2 error. Next, why remove mean cortical thickness or FDG across the whole brain? Unlike brain volume, which can vary as a function of head size (i.e., intracranial volume), cortical thickness does not suffer from this limitation. There is no need, so far as I know, for correcting for mean cortical thickness. Similarly, for FDG PET, if the exact MBq and other specifics were known for each participant, which they should be in this design, then the Distribution Volume Ratio (DVR) can be calculated using Logan graph analysis. Standardized Uptake Volume Ratio (SUVR) can be calculated if need be, but in that case it is calculated by taking the mean ROI uptake value of a given region (e.g., medial temporal gyrus) and relativizing it to a reference region (e.g., cerebellum). Regressing out the global FDG mean is not standard practice. Indeed, given that age leads to hypometabolism or atrophy in a front-to-back specific manner, and that AD and related dementias have a specific pattern of atrophy, relativizing to the global mean is not recommended; it will introduce bias in ROIs and skew results. Finally, for the statistics section, it would be prudent to mention the FDR correction and other statistical concerns here instead of in the rest of methods.

For results, first, I am confused by the delineation of the middle-aged and aged groups. The groups appear to have been split along the mean or median age for segregation. However, the middle-aged group includes young adults, whereas the aged group has late middle-aged adults. In other words, the “middle-aged” group is actually young adult + middle-aged adults and the “old” or aged group is actually middle-aged + early aged adults. This will lead to bias in results, and may undermine the authors’ claim that Supplementary Figure 1 and Figure 2 results suggesting discordancy may, in fact, be due to methodological limitations. Next, for Figure 2, it is unclear which brain regions were chosen for Figure 2A-C. The axes should be relabeled or the information included in the main text. As a minor point, what is the value for the color bars? In the Figure 2 legend, the y-axis is defined at PET values to the 3^{rd} power, but I am unfamiliar with this approach and am unsure why it was done. Next, for network similarity, I am confused by the P(existence|connected) and P(inexistence|connected) results. If I am not mistaken, the mean P value (?) for the “middle-aged” and “aged” groups are non-significant. Further, it is not unexpected that permutation testing would reveal a significant difference between the “middle-aged” and “aged” groups. I may have missed it, but when was permutation testing described in the methods section? Further, I am confused by contrasting P values between FC and SC that are described using an FDR P value. Instead, should not Spearman correlation coefficients be used? At present, I do not agree with the claim that FC and SC show different patterns of results. Next, Figure 3 seems odd to me. As a function of aging, there is atrophy and progressive hypometabolism in a well defined series of regions. In the “middle-aged” group, connectivity likeness is found in post-central gyrus, cuneus, and portions of lateral parietal lobe. One would expect a stronger FC and SC connectivity earlier in life. In the old age group, by contrast, connectivity likeness is found throughout most of the brain. This makes sense in areas which show both hypometabolism and atrophy starting in the 2^{nd} decade and becoming shallower after the 5^{th} decade (e.g., prefrontal cortex), but then there are areas of significance which show shallow to non-existent changes in normal aging (e.g., post-central gyrus). I am not sure why older adults would show stronger correlations between FC and SC in regions that are not traditionally affected by aging, or if so are affected modestly. Next, I am unsure why Figure 4 is present in the manuscript. By definition, nodes and vectors representing FC and SC will vary as a function of normal aging. That FC is different from SC is not surprising. For example, the Default Mode Network is perhaps the most important functional network in the brain, and is required for corroboration between regions related to self-reflection and

long-term memory (i.e., posteromedial parietal regions) and areas related to judgment and higher-order cognition to order self-referential memory logically (i.e., medial prefrontal regions). While tracts like Superior Longitudinal Fasciculus and others provide some connection between the areas, functionally there is strong connectivity. I fail to see, then, why modest comparisons between “middle-aged” and “aged” groups are shown, because we know these differences exist and that they vary over the lifespan.

For discussion, I again have concerns. In the first paragraph, it is stated that FC and SC discordancy was larger during aging. Yet, for Figure 3, the “aged” group is described as showing “significant higher SC-FC coupling. These statements seem to be in conflict. Next, it is indicated that surface parameters from FreeSurfer were not correlated between imaging modalities types, an that cortical thickness was focused on. I was not aware of this non-significance while reading the results section. Some revision may be required to make this more clear to the reader. In particular, clarifying what are regional vs. connective level analyses may be helpful, as that is not self-evident in the text. Next, while I agree with the statement that “age affects CTh and brain glucose metabolism in different ways,” I am not sure that is adequately shown in results and that the differences between the two age groups seem to be paradoxical (e.g., Figure 3 showing stronger FC and SC correlations in the older vs. younger group). Next, the claim is put forward that because FC and SC correlations were stronger in aging, it reflects “poorer cognitive performance and decreasing awareness/consciousness.” While studies are cited to support this claim, no behavioral data is presented in the report. It is, at best, tenuous to make a conclusion about behavior or neural underpinnings of behavior when no such data is presented. Lastly, I appreciate the limitations section mentioning the age range. I think several other limitations could also be added, some of which are described above.