Test-retest reproducibility of structural and proxy estimates of brain connectivity at rest

Aldana Lizarraga , Arianna Sala , Kathrin Koch , Igor Yakushev

PII: S1053-8119(25)00561-0
DOI: https://doi.org/10.1016/j.neuroimage.2025.121558
Reference: YNIMG 121558

To appear in: *NeuroImage*

Received date: 15 March 2025
Revised date: 25 July 2025
Accepted date: 24 October 2025

Please cite this article as: Aldana Lizarraga , Arianna Sala , Kathrin Koch , Igor Yakushev , Test-retest reproducibility of structural and proxy estimates of brain connectivity at rest, *NeuroImage* (2025), doi: https://doi.org/10.1016/j.neuroimage.2025.121558

Highlights

- Across all estimates stronger connections exhibit highest test-retest reproducibility
- Structural connectivity is the most reproducible estimate of brain connectivity
- Intersubject covariance of regional 18F-fluorodeoxyglucose uptake tops proxy estimates
- Thresholding boosts connection reproducibility

Test-retest reproducibility of structural and proxy estimates of brain connectivity at rest

Aldana Lizarraga [a]*, Arianna Sala [a,b]*, Kathrin Koch [c], Igor Yakushev [a]

* *equally contributed*

[a] Department of Nuclear Medicine, School of Medicine, Technical University of Munich, 81675 Munich, Germany

[b] Coma Science Group, GIGA Consciousness, University of Liege and Centre du Cerveau2, University Hospital of Liege, Liege 4000, Belgium

[c] Department of Neuroradiology, School of Medicine, Klinikum Rechts der Isar, Technical University of Munich, 81675 Munich, Germany

Corresponding author:

Igor Yakushev, MD
Dept. of Nuclear Medicine
Technical University of Munich
Ismaninger Str. 22
81675 Munich, Germany
Phone: +49 (0) 89-4140-6085
Fax: +49 (0) 89-4140-7431
Email: igor.yakushev@tum.de

Abstract

While structural connectivity (SC) reflects actual neural connections, proxy estimates of brain connectivity such as functional connectivity (FC) from functional MRI, intersubject covariance of regional gray matter volume (GMV_{cov}) from structural MRI, and intersubject covariance of regional 18F-fluorodeoxyglucose uptake (FDG_{cov}) from PET are derived from statistical dependencies between regional measurements. Understanding reproducibility of these estimates is crucial for capturing physiological and pathological changes in brain connectivity.

Here, we compared test-retest reproducibility of group-level SC, FC, GMV_{cov}, and FDG_{cov} in the same 55 healthy subjects at rest using a simultaneous PET/MRI acquisition protocol. Sparsity-based thresholding resulted in coefficients of variation of 2.7%, 3.1%, 3.6%, and 5.1% for SC, FDG_{cov}, GMV_{cov}, and FC, respectively. Strength-based thresholding resulted in relative proportion of repeatedly (test-retest) present connections (PRPC) of 77%, 56%, and 54% for GMV_{cov}, FDG_{cov}, and FC, respectively. Absolute PRPC were 2.50%, 0.64%, and 0.25% for FDG_{cov}, FC, and GMV_{cov}, respectively.

Overall, stronger connections exhibited higher reproducibility, arguing in favour of thresholding as a necessary analytical step. SC was most reproducible. Among the proxy estimates, the highest absolute PRPC was found for FDG_{cov}. FDG_{cov} enables to study brain connectivity in a reproducible manner over the largest part of the brain.

Key words: positron emission tomography, FDG PET, molecular connectivity, networks, functional connectivity

1. Introduction

Connectomics studies complex interactions between different elements of the brain. This branch of neuroscience has expanded incredibly in the last decades, increasing our knowledge on the role of brain networks in normal and pathological cognition. Herewith, magnetic resonance imaging (MRI) techniques have made the major contribution (Jbabdi et al., 2015). Structural connectivity (SC) refers to anatomical white matter pathways connecting gray matter regions. It is commonly estimated through 3D virtual reconstructions of white matter (WM) tracts from MRI diffusion weighted imaging (DWI) data (Smith et al., 2022). Functional connectivity (FC) refers to functional interactions between brain regions and is an indirect proxy of neural activity (Drew, 2019). It is estimated as statistical dependence of regional blood oxygen level dependent (BOLD) signal from functional MRI (fMRI) data. Moreover, connectivity approaches have been successfully applied to imaging modalities, where typically only one image per subject is available. In a relatively homogenous sample, these applications allow to derive measures of a shared covariance that can be considered as population-invariant part of the connectome (Sala et al., 2022, 2023; Sporns, 2010). Thus, MRI T1-weighted images have been widely used to derive a shared covariance of morphologic features across subjects as index of brain connectivity (Seghier and Price, 2018; Wang and He, 2024). It was interpreted in terms of shared trophic influences mediated by anatomical connections and neural plasticity (Romero-Garcia et al., 2018). Similarly, regional glucose consumption from 18F-fluorodeoxyglucose (FDG) positron emission tomography (PET) data, a direct proxy of neural activity (Stoessl, 2017), has been analyzed in the context of brain connectivity, also known as “metabolic connectivity” (Yakushev et al., 2017). It is commonly interpreted as functional interactions between brain regions (Horwitz et al., 1984). In the last years, intersubject estimation of brain connectivity from FDG-PET data has arisen as a valuable approach (Sala et al., 2023), helping to understand mechanisms of healthy aging (Mertens et al., 2022), neuropsychiatric diseases (Nicastro et al., 2021) and to improve diagnostic accuracy (Perovnik et al., 2023). While SC reflects actual anatomical connectivity and is derived at a single subject level, proxy estimates of (non-anatomical) brain connectivity such as FC, intersubject covariance of regional GM volume (GMV_{cov}) and of regional FDG uptake (FDG_{cov}) are derived from statistical dependencies between regional measurements, typically at the group level (Lizarraga et al., 2023).

Assessment of test-retest reproducibility of a measure is essential to design properly powered studies and assign a degree of confidence to the inference derived from it (Zuo et al., 2019). Importantly, reproducibility error sets a limit to the minimal amount of change in a measured signal that can be associated with the true effect of a certain task or disease and is not caused by spurious variability due to technical or biological noise (Zuo et al., 2019). Previous studies found SC to be more reproducible than FC (Osmanlıoğlu et al., 2020; Welton et al., 2015). Whereas rare studies

have assessed reproducibility of GMV_{cov} (Carmon et al., 2020; Geerligs et al., 2016), it has not been compared with reproducibility of SC and FC so far. Moreover, no data exist on reproducibility of FDG_{cov}, despite its increasing popularity.

This study for the first time compared reproducibility of SC, FC, FDG_{cov}, and GMV_{cov} in the same subjects and using the same experimental set-up. To this end, MRI and FDG-PET data were acquired simultaneously in a large group of healthy individuals on a hybrid PET/MR scanner twice.

2. Methods and materials

2.1 Subjects

The study was approved by the Federal Office for Radiation Protection and the Ethics Committee of the TUM University Hospital (project number 399/13). Written, informed consent was obtained from all participants. The healthy subjects were scanned in a 3T PET/MR Siemens Biograph mMR scanner twice, with a time interval of eight weeks. Because no differences between two identical acquisitions were found (Ripp et al., 2022), we treated these data as test-retest. The subjects were free of cognitive deficits, neuropsychiatric diseases, and contraindications for MRI. The data of 55 participants (24 females, age 57±4 years) were available for the present study; they have been used in other studies previously (Lizarraga et al., 2023; Ripp et al., 2022).

2.2 Data acquisition

All data for the present study were acquired under resting conditions. After injecting of approximately 100 MBq FDG intravenously the participants relaxed in a quiet room, while PET/MR scanning began 30 min later. T1-weighted images of 1.0×1.0×1.0 mm^3 were acquired with a magnetization-prepared rapid gradient-echo (MP-RAGE) sequence. DWI of 2.0×2.0×2.0 mm^3 was performed using a single-shot echo-planar imaging (ss-EPI) sequence with 30 diffusion directions with a b value of 800 s/mm2 and one volume with b=0 s/mm2. For the rs-fMRI acquisition participants were instructed to close their eyes and think of nothing in particular. 212 volumes were acquired using a Prospective Acquisition Correction EPI (PACE) sequence with 3.0x3.0x3.0 mm^3. Finally, an ultra-short echo time (UTE) μ-map sequence was acquired for attenuation correction, and a dual echo gradient echo sequence (GRE) was included to estimate the field map and correct the images for susceptibility induced distortions. For a more detailed description of the imaging protocol see supplementary material of our previous study (Ripp et al., 2022).

2.3 Grey matter parcellation

We parcelled the GM of each subject into 106 regions. To this end, Automated Anatomical Labeling 2 (AAL2) atlas (Rolls et al., 2015), which includes cortex, subcortex and cerebellum, was

non-linearly transformed from the MNI152 T1 space to each individual T1 space using Advanced Normalization Tools (ANTs) (Avants et al., 2011). Additionally, T1 images were segmented into different tissues classes using Statistical Parametric Mapping (SPM) segmentation tool. After that, GM probability maps were thresholded at 0.5, binarized, and segmented into the regions (Yakushev et al., 2022). Certain regions of the original parcellation were combined or removed in order to minimize partial volume effect in PET images; for details on this step, as well as the definitive list of regions, see supplementary material in Lizarraga et al. (2023).

2.4 SC

We obtained SC networks from DWI data using tools from FMRIB Software Library (FSL). After a visual inspection of the images with FSLeyes, we corrected for susceptibility induced distortions, eddy currents, inter-volume movement and signal dropout using *eddy* (Andersson and Sotiropoulos, 2016). Then, we extracted the brain tissue with BET (Smith, 2002) and fit the Ball and Sticks diffusion model with N=2 using BedpostX (Behrens et al., 2007, 2003). After that, we performed whole brain probabilistic tractography using a WM seeding approach with ProbtrackX. For that, 0.5 thresholded masks of GM, WM, and CSF were used as target, seed and exclusion masks, respectively. Although the tractography ran in the DWI space, we obtained the outputs in the T1 space. As a measure of SC we used the number of streamlines connecting two GM regions and normalized them by the surface area of those regions in the WM-GM interface to compensate for surface-driven effects on streamline counts (Bonilha et al., 2015; van den Heuvel and Sporns, 2011). In order to construct a group SC network, we calculated an average across subjects.

2.5 FC

The diagnostic tool tsdiffana was utilized in MATLAB software (MATLAB - MathWorks - MATLAB & Simulink) to detect potential problems in rs-fMRI datasets. After removal of the first three volumes of the BOLD series, the images were corrected for slice timing and realigned to the volume acquired temporally in the middle using SPM12 (https://www.fil.ion.ucl.ac.uk/spm/software/spm12/) tools. We created WM and CSF masks applying a threshold of 0.99 to the probability maps generated before. Those masks, as well as the GM parcellation, were linearly transformed from the T1 space to the BOLD space using the FSL registration tool FLIRT with a nearest neighbor interpolation (Jenkinson et al., 2002). After that, we regressed the GM and CSF signal components using *fsl_glm* and applied a temporal band-pass filter (0.009-0.08 Hz) with MATLAB. Next, we extracted the signal from each region and calculated the Pearson correlation between each pair of regions for each subject. We chose the Pearson correlation as a popular metric in the field of brain connectivity (Mahadevan et al., 2021; Sala et al., 2023) and for a direct comparison with GMV_{cov} and FDG_{cov} (see below). In order to build a group FC network, we applied a Fisher z-transformation to the correlation coefficients,

averaged them across subjects and transformed them back to Pearson correlation coefficients (Di et al., 2017; Lizarraga et al., 2023).

2.6 GMV_{cov}

A GMV_{cov} network was constructed from T1-weighted images. We decided to extract volume rather than thickness a) to keep cerebellum and subcortical structures in the analyses (Lizarraga et al., 2023) and b) due to higher reproducibility of volume-based as compared to thickness-based networks (Carmon et al., 2020). After GM parcellation of T1-weighted images as described above, regional GMV was calculated for each region as a sum of GM probabilities multiplied by the voxel size. After that, we regressed out the total GMV of each subject (Wu et al., 2012). Finally, Pearson correlation of the residuals between pairs of regions was calculated across subjects.

2.7 FDG_{cov}

Using Siemens e7 Tools (Siemens Molecular Imaging, Knoxville, USA), we performed an offline reconstruction of a single FDG-PET summation image from the first 30 min of acquisition (30 to 60 min post injection) for each subject; for technical details see (Ripp et al., 2021). After a linear coregistration of PET images to the individual T1 space with SPM, FDG uptake was extracted from each region and normalized by the total GM uptake. Finally, FDG_{cov} was calculated as Pearson correlations across subjects.

2.8 Test-retest reproducibility

Selection of metrics for comparison of test-retest reproducibility of connectivity estimates from four neuroimaging techniques is challenging. First, we computed the rank Spearman correlation coefficient (SCC) between test and retest connections with all possible connections in the unthresholded matrix with their original signs. This common reproducibility metric (Geerligs et al., 2016; Osmanlıoğlu et al., 2020) measures how well the correspondence between test and retest network edges can be described using a monotonic function. SCC has an easy interpretation and is less sensitive to outliers and changes in the spread of data than the Pearson's correlation coefficient.

Because reproducibility can vary significantly across connections within a network (Tsai, 2018; Zhao et al., 2015), we in addition utilized the coefficient of variation (CV) between test and retest connections with all possible connections in the unthresholded matrix with their original signs (matrices can be found in Figure 1 in Lizarraga et al., 2023). CV is a statistical measure of the dispersion of data points around the mean that has also been extensively used in reproducibility studies (Chou et al., 2012; Tsai, 2018; Yuan et al., 2019). CV standardizes variability by expressing it as a percentage of the mean. The lower the value of CV, the more precise the estimate. CV of each connection strength in the group adjacency matrices was calculated according to the formula

$$CV = 100\% \frac{std\ (test, retest)}{mean\ (test, retest)}$$

Because the sign of CV is not meaningful in the reproducibility analyses, the absolute value of CV (abs(CV)) is reported (Biswal et al., 2010).

As reproducibility depends on the connection strength (Geerligs et al., 2016; Shehzad et al., 2009), we in addition compared reproducibility of 8% strongest connections for each estimate after applying a 92% sparsity threshold to all four networks. This threshold was selected after inspecting the percentage of significant connections for proxy estimates (8% for FC, 14% for GMV_{cov} and 45% for FDG_{cov}) and to include only significant connections (p<0.05) for all of them. Thus, we computed median abs(CV) in the range 92-100%. Of note, to calculate CV correctly, a connection has to be present in both thresholded networks, so the sparsity used to report the results is not the one used to threshold the networks, but the sparsity calculated when including the commonly present connections after thresholding.

Finally, we compared reproducibility after applying an absolute threshold for connectivity strength, a frequent practice in connectomics (Osmanlıoğlu et al., 2020; Tozzi et al., 2020). Since structural connections have a different nature and range, SC was not included in this analysis. To quantify reproducibility of resultant binary networks, we calculated the proportion of repeatedly present connections (PRPC) as those connections that remained in test and retest sessions after thresholding. Afterwards, we applied a threshold of r=0.5, keeping only connections with absolute value greater than 0.5 (Mukaka, 2012). The absolute proportion of repeatedly present connections (aPRPC) was defined as percentage of repeatedly present connections (N_{RPC}) to the total number of potential connections in a network.

$$aPRPC = 100\% \frac{N_{RPC}}{\frac{N_{ROIs} \times (N_{ROIs} - 1)}{2}}$$

To account for the different number of present connections among the techniques, we computed the percentage of N_{RPC} in relation to the union, i.e., connections in test *or* retest sessions, hereafter called relative proportion of repeatedly present connections (rPRPC). It is equivalent to a Dice Coefficient between test and retest binary networks.

$$rPRPC = 100\% \frac{N_{RPC}}{N_{test \cup retest}}$$

aPRPC and rPRPC provide complementary information. While aPRPC indicates the extent to which the connectome can be reproducibly derived at a given threshold, rPRPC shows the spatial overlap between the test and retest connectomes. In addition to the 0.5 cut-off above, we conducted sensitivity analyses using thresholds for absolute connection values above 0.5, evenly spaced

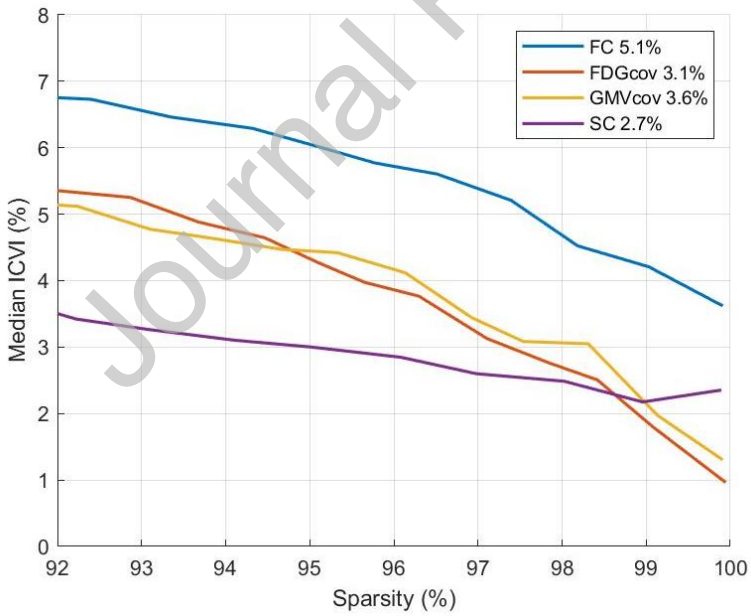
between 0.50 to 0.99, as well as across the entire range from 0 to 1. A representative value was calculated for each connectivity estimate as AUC normalized by the corresponding range of thresholds.

3. Results

SCC between the test and retest connection strengths in the unthresholded group networks are shown in Figure S1. All SCC were strong and statistically significant (p<0.05, Bonferroni corrected): 0.98 for SC, followed by 0.96 for FC, 0.95 for FDG_{cov}, and 0.93 for GMV_{cov}. Across all 4 estimates, weaker connections showed a higher variability between test and retest.

CV was calculated for each potential connection in the group unthresholded adjacency matrices (Figure S2). Considering all potential connections, no significant (p>0.05) differences were found between median CVs of the estimates: 23±20 for SC, 20±15 for FC, 22±16 for FDG_{cov}, and 24±17 for GMV_{cov} (Figure S3). However, when thresholding the adjacency matrices to keep only the strongest connections, significant differences between the estimates emerge (Figure 1): SC showed the lowest CV (2.7% on average), followed by FDG_{cov} (3.1%), GMV_{cov} (3.6%), and FC (5.1%).

Figure 1. Reproducibility as function of sparsity



Median absolute value of CVs for each estimate as function of effective sparsity. The box indicates average values over the sparsity range.

For proxy estimates, the percentages of connections in each group, from negligible to very high (Mukaka, 2012), are presented in Table 1. The maximum amount of very high connections was found for FDG_{cov}, the minimum for FC. GMV_{cov} presented the highest, while FDG_{cov} the lowest number of negligible connections.

Table 1. Categorization of proxy estimates of brain connectivity according to strength (Pearson's correlation coefficients)

	FC, %		FDG_{cov}, %		GMV_{cov}, %	
	test	retest	test	retest	test	retest
Very high (0.9 – 1.0)	0.00	0.04	0.36	0.31	0.04	0.02
High (0.7 – 0.9)	0.66	0.83	2.93	1.64	0.25	0.23
Moderate (0.5 – 0.7)	2.01	3.23	14.90	10.93	0.81	0.86
Low (0.3 – 0.5)	11.27	13.48	26.74	25.10	8.21	8.19
Negligible (0.0 – 0.3)	86.06	82.43	55.08	62.03	90.69	90.69

aPCRP at the absolute threshold of r=0.5 are presented in Table 2 and Figure 2. Herewith, connections in the categories negligible and low were (Table 1) treated as "absent", whereas the remaining categories were treated as "present" connections. FDG_{cov} showed the highest aPRPC, followed by FC and GMV_{cov}. Figure 3 shows aPRPC and rPRPC calculated for the thresholds above 0.5. Overall, FDG_{cov} exhibited the highest aPRPC (2.5%), followed by FC (0.64%), and GMV_{cov} (0.25%). GMV_{cov} presented the highest rPRPC (77%), followed by FDG_{cov} (56%), and FC (54%). Similar results were obtained for the full range of absolute values of connection strengths (0-1) (Figure S4).

Table 2. Snapshot of retained connections after 0.5 absolute thresholding

	FC, %	FDG_{cov}, %	GMV_{cov}, %
Repeatedly present (aPCRP)	2.6	11.5	0.9
Repeatedly absent	95.8	80.4	98.7
Present in test	2.7	18.2	1.1
Present in retest	4.1	12.9	1.1

Figure 2. Snapshot of repeatedly present connections after 0.5 absolute thresholding

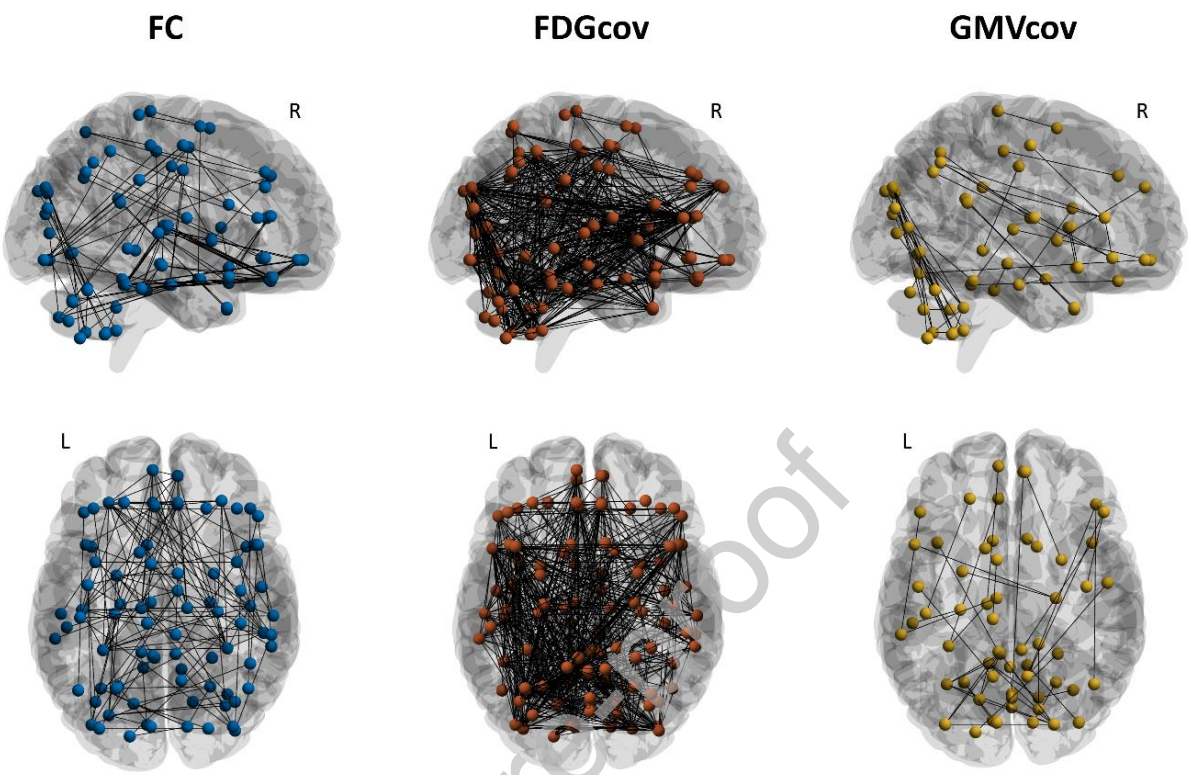
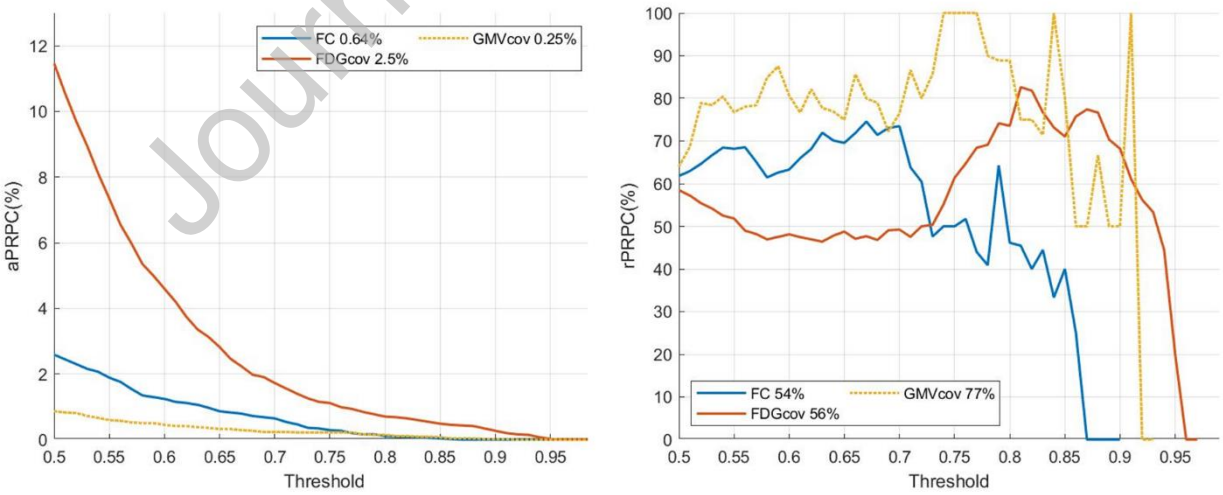


Figure 3. aPRPC and rPRPC as a function of the absolute threshold



FC (blue), FDG_{cov} (orange), GMV_{cov} (yellow). Values indicate the average over the sparsity range.

4. Discussion

In this study we compared test-retest reproducibility of SC, FC, FDG_{cov} and GMV_{cov} in the same healthy subjects and using the same experimental set-up. Across all estimates stronger connections exhibited higher reproducibility. Overall, SC was most reproducible. Among the proxy estimates of brain connectivity, the lowest CV was found for FDG_{cov}. While the relative proportion of reproducible, strong connections was similar for FDG_{cov} and FC, the highest absolute proportion of reproducible, strong connections was found for FDG_{cov}. Thus, FDG_{cov} enables to explore brain connectivity in a reproducible manner over a substantially larger part of the brain than FC and GMV_{cov}.

Across all estimates (SC, FC, FDG_{cov} and GMV_{cov}) stronger connections throughout the brain exhibited higher reproducibility. That is, when connections approach the extremes of their distributions, the scatter plots become narrower, resembling the identity function (Figure S1). Furthermore, all estimates exhibited a monotonically decreasing CV in parallel with increasing sparsity levels, with a more pronounced negative slope for SC. Thus, stronger connections have a lower CV, that is higher reproducibility. Similar observations have been previously reported for SC (Tsai, 2018) and FC (Geerligs et al., 2016), whereas no data on FDG_{cov} and GMV_{cov} have existed so far. Overall, our findings indicate that removing weaker connections through thresholding increases reproducibility of all studied estimates of brain connectivity.

Interestingly, stronger connections (mainly intralobe and homotopic interhemispheric) are also those that show a higher similarity across these four connectivity estimates (Lizarraga et al., 2023), indirectly arguing in favor of their true rather than false-positive nature. When no thresholding was applied, SCC, a common index of reproducibility, was found to be high (>0.9) for all estimates. However, further analyses revealed that the absolute majority of the data points (82 to 99%) reflected connections of low to negligible strength, i.e., likely noise. Thus, SCC has a limited applicability in the setting of non-thresholded matrices.

SC was previously found to have good reproducibility (Bonilha et al., 2015; Messaritaki et al., 2019; Tsai, 2018). Moreover, structural connectomes are more reproducible than functional ones (Chamberland et al., 2017; Lin et al., 2015; Osmanlıoğlu et al., 2020; Welton et al., 2015). These results have been explained by the dynamic nature of FC (Liégeois et al., 2019). Our study confirms these previous findings. Moreover, we found SC to be also more reproducible than FDG_{cov} and GMV_{cov}.

Only two studies have assessed reproducibility of covariance networks derived from morphological features (Carmon et al., 2020; Geerligs et al., 2016). Carmon et al. (2020) found cortical thickness to be associated with lower reproducibility than volume and surface area. Unfortunately, their results cannot be compared to ours, as they used a different measure to assess reproducibility. Similarly, Geerligs et al. (2016) found that GMV_{cov} networks exhibited the highest

reproducibility when using the AAL atlas, which has the lowest resolution among the tested parcellations. They reported a SCC above 0.8, which is slightly lower than our results. These two prior studies support the methodological choices in our study, i.e., GM volume as morphological feature and the AAL2 atlas as parcellation.

Among the proxy estimates of brain connectivity, FDG_{cov} exhibited the highest aPRPC across any absolute threshold, followed by FC and GMV_{cov}. These findings suggest that FDG_{cov} enables a reproducible exploration of brain connectivity across a substantially larger portion of the brain than other proxy estimates of brain connectivity. Specifically, FDG_{cov} "catches" roughly 4 times more reproducible connections than FC does. The lower reproducibility of FC can be well explained by the very dynamic character of the BOLD signal and a higher contribution of non-neural sources to the signal (Korponay et al., 2024). Of note, GMV_{cov} demonstrated higher rPRPC than both FC and FDG_{cov}. This means that GMV_{cov} network has, on average, the largest spatial overlap between test and retest sessions. This can be explained by a higher reproducibility of GMV compared to FDG uptake and BOLD signal, because GMV is a (static) feature of brain structure rather than a (dynamic) feature of brain function. Consequently, strong GMV_{cov} connections are highly reproducible. However, the proportion of these connections is very small, below 1% of all potential connections. Thus, GMV_{cov} allows for (reproducible) exploration of only a small fraction of the brain connectome.

Whereas our results clearly argue for necessity of thesholding of the connectivity estimates, there is no one right level of the threshold. It would primarily depend on the objective of a given study. Specifically, if the study focuses on a connectivity pattern at the whole brain level, a more liberal Pearson's correlation threshold of 0.5 can be acceptable. However, if the focus is on the anatomical/spatial precision, e.g., *which* connections a specific region or a network have, at least 0.8 spatial overlap (rPRPC) should apply. In this situation, a Pearson's correlation coefficient of 0.8 and 0.7 should be used as threshold for FDG_{cov} and GMV_{cov}, respectively, while FC marginally does not reach this value (figure 3).

Whether the current findings apply to connectomes of clinical populations remains to be determined. To the best of our knowledge, some previous clinical studies have assessed test-retest reproducibility of *single* estimates of brain connectivity (Camp et al., 2024; Cole et al., 2014; Ma et al., 2007; Somandepalli et al., 2015), with similar results for healthy subjects and clinical populations being found for FDG_{cov} (Ma et al., 2007) and GMV_{cov} (Cole et al., 2014), but mixed findings for FC (Camp et al., 2024; Somandepalli et al., 2015). Furthermore, test-retest reproducibility is likely to vary with e.g., subject age, acquisition and processing parameters, an important topic for future studies. Of note, apart from these general factors some estimates may be susceptible to specific influences such as depletion and activation of microglia for FDG_{cov} (Gnörich et al., 2023).

In conclusion, this study found that irrespective of the brain connectivity estimate, stronger connections exhibited higher reproducibility. Overall, SC was most reproducible. Among the proxy estimates of brain connectivity, FDG_{cov} exhibited the highest reproducibility in terms of CV and aPRPC. Thus, FDG_{cov} enables to explore brain connectivity in a reproducible manner over a substantially larger part of the brain than FC and GMV_{cov}. Ultimately, apart from reproducibility the choice of the most appropriate connectivity estimate for a study should be made under consideration of study goals, hypothesis, available data, and methodological expertise.

Acknowledgment

This study was partially funded by the German Research Foundation (DFG), grant to IY [grant number YA 373/3–1] and KK [grant number KO 3744/8–1].

Author contributions

AL, AS, and IY contributed to the conception and design of the study. AL analyzed the data. AL, AS and IY wrote the manuscript. All authors contributed to the manuscript revision and approved the submitted version.

Conflict of interests

The authors declare that they have no conflicts of interest related to this study.

Code availability

Code is available per reasonable request.

References

Andersson, J.L.R., Sotiropoulos, S.N., 2016. An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. NeuroImage 125, 1063–1078. https://doi.org/10.1016/j.neuroimage.2015.10.019

Avants, B.B., Tustison, N.J., Song, G., Cook, P.A., Klein, A., Gee, J.C., 2011. A reproducible evaluation of ANTs similarity metric performance in brain image registration. NeuroImage 54, 2033–2044. https://doi.org/10.1016/j.neuroimage.2010.09.025

Behrens, T.E.J., Berg, H.J., Jbabdi, S., Rushworth, M.F.S., Woolrich, M.W., 2007. Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? NeuroImage 34, 144–155. https://doi.org/10.1016/j.neuroimage.2006.09.018

Behrens, T.E.J., Woolrich, M.W., Jenkinson, M., Johansen-Berg, H., Nunes, R.G., Clare, S., Matthews, P.M., Brady, J.M., Smith, S.M., 2003. Characterization and propagation of uncertainty in diffusion-weighted MR imaging. Magn. Reson. Med. 50, 1077–1088. https://doi.org/10.1002/mrm.10609

Biswal, B.B., Mennes, M., Zuo, X.-N., Gohel, S., Kelly, C., Smith, S.M., Beckmann, C.F., Adelstein, J.S., Buckner, R.L., Colcombe, S., Dogonowski, A.-M., Ernst, M., Fair, D., Hampson, M., Hoptman, M.J., Hyde, J.S., Kiviniemi, V.J., Kötter, R., Li, S.-J., Lin, C.-P., Lowe, M.J., Mackay, C., Madden, D.J., Madsen, K.H., Margulies, D.S., Mayberg, H.S., McMahon, K., Monk, C.S., Mostofsky, S.H., Nagel, B.J., Pekar, J.J., Peltier, S.J., Petersen, S.E., Riedl, V., Rombouts, S.A.R.B., Rypma, B., Schlaggar, B.L., Schmidt, S., Seidler, R.D., Siegle, G.J., Sorg, C., Teng, G.-J., Veijola, J., Villringer, A., Walter, M., Wang, L., Weng, X.-C., Whitfield-Gabrieli, S., Williamson, P., Windischberger, C., Zang, Y.-F., Zhang, H.-Y., Castellanos, F.X., Milham, M.P., 2010. Toward discovery science of human brain function. Proc. Natl. Acad. Sci. U.S.A. 107, 4734–4739. https://doi.org/10.1073/pnas.0911855107

Bonilha, L., Gleichgerrcht, E., Fridriksson, J., Rorden, C., Breedlove, J.L., Nesland, T., Paulus, W., Helms, G., Focke, N.K., 2015. Reproducibility of the Structural Brain Connectome Derived from Diffusion Tensor Imaging. PLoS ONE 10, e0135247. https://doi.org/10.1371/journal.pone.0135247

Camp, C.C., Noble, S., Scheinost, D., Stringaris, A., Nielson, D.M., 2024. Test-Retest Reliability of Functional Connectivity in Adolescents With Depression. Biol Psychiatry Cogn Neurosci Neuroimaging 9, 21–29. https://doi.org/10.1016/j.bpsc.2023.09.002

Carmon, J., Heege, J., Necus, J.H., Owen, T.W., Pipa, G., Kaiser, M., Taylor, P.N., Wang, Y., 2020. Reliability and comparability of human brain structural covariance networks. NeuroImage 220, 117104. https://doi.org/10.1016/j.neuroimage.2020.117104

Chamberland, M., Girard, G., Bernier, M., Fortin, D., Descoteaux, M., Whittingstall, K., 2017. On the Origin of Individual Functional Connectivity Variability: The Role of White Matter Architecture. Brain Connectivity 7, 491–503. https://doi.org/10.1089/brain.2017.0539

Chou, Y. -h., Panych, L.P., Dickey, C.C., Petrella, J.R., Chen, N. -k., 2012. Investigation of Long-Term Reproducibility of Intrinsic Connectivity Network Mapping: A Resting-State fMRI Study. AJNR Am J Neuroradiol 33, 833–838. https://doi.org/10.3174/ajnr.A2894

Cole, J.H., Farmer, R.E., Rees, E.M., Johnson, H.J., Frost, C., Scahill, R.I., Hobbs, N.Z., 2014. Test-Retest Reliability of Diffusion Tensor Imaging in Huntington's Disease. PLoS Curr 6, ecurrents. https://doi.org/10.1371/currents.hd.f19ef63fff962f5cd9c0e88f4844f43b

Di X, Gohel S, Thielcke A, Wehrl HF, Biswal BB. Do all roads lead to Rome? A comparison of brain networks derived from inter-subject volumetric and metabolic covariance and moment-to-moment hemodynamic correlations in old individuals. Brain Struct Funct. 2017; 222(8): 3833-3845. doi: 10.1007/s00429-017-1438-7.

Drew, P.J., 2019. Vascular and neural basis of the BOLD signal. Curr Opin Neurobiol 58, 61–69. https://doi.org/10.1016/j.conb.2019.06.004

Geerligs, L., Cam-CAN, Henson, R.N., 2016. Functional connectivity and structural covariance between regions of interest can be measured more accurately using multivariate distance correlation. NeuroImage 135, 16–31. https://doi.org/10.1016/j.neuroimage.2016.04.047

Gnörich, J., Reifschneider, A., Wind, K., Zatcepin, A., Kunte, S.T., Beumers, P., Bartos, L.M., Wiedemann, T., Grosch, M., Xiang, X., Fard, M.K., Ruch, F., Werner, G., Koehler, M., Slemann, L., Hummel, S., Briel, N., Blume, T., Shi, Y., Biechele, G., Beyer, L., Eckenweber, F., Scheifele, M., Bartenstein, P., Albert, N.L., Herms, J., Tahirovic, S., Haass, C., Capell, A., Ziegler, S., Brendel, M., 2023. Depletion and activation of microglia impact metabolic connectivity of the mouse brain. J Neuroinflammation 20, 47. https://doi.org/10.1186/s12974-023-02735-8

Horwitz, B., Duara, R., Rapoport, S.I., 1984. Intercorrelations of Glucose Metabolic Rates between Brain Regions: Application to Healthy Males in a State of Reduced Sensory Input. J Cereb Blood Flow Metab 4, 484–499. https://doi.org/10.1038/jcbfm.1984.73

Jbabdi, S., Sotiropoulos, S.N., Haber, S.N., Van Essen, D.C., Behrens, T.E., 2015. Measuring macroscopic brain connections in vivo. Nat Neurosci 18, 1546–1555. https://doi.org/10.1038/nn.4134

Jenkinson, M., Bannister, P., Brady, M., Smith, S., 2002. Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. NeuroImage 17, 825–841. https://doi.org/10.1006/nimg.2002.1132

Korponay C, Janes AC, Frederick BB. Brain-wide functional connectivity artifactually inflates throughout functional magnetic resonance imaging scans. Nat Hum Behav. 2024; 8(8): 1568-1580. doi: 10.1038/s41562-024-01908-6.

Liégeois, R., Li, J., Kong, R., Orban, C., Van De Ville, D., Ge, T., Sabuncu, M.R., Yeo, B.T.T., 2019. Resting brain dynamics at different timescales capture distinct aspects of human behavior. Nat Commun 10, 2317. https://doi.org/10.1038/s41467-019-10317-7

Lin, Q., Dai, Z., Xia, M., Han, Z., Huang, R., Gong, G., Liu, C., Bi, Y., He, Y., 2015. A connectivity-based test-retest dataset of multi-modal magnetic resonance imaging in young healthy adults. Sci Data 2, 150056. https://doi.org/10.1038/sdata.2015.56

Lizarraga, A., Ripp, I., Sala, A., Shi, K., Düring, M., Koch, K., Yakushev, I., 2023. Similarity between structural and proxy estimates of brain connectivity. J Cereb Blood Flow Metab 0271678X231204769. https://doi.org/10.1177/0271678X231204769

Ma, Y., Tang, C., Spetsieris, P.G., Dhawan, V., Eidelberg, D., 2007. Abnormal metabolic network activity in Parkinson's disease: test-retest reproducibility. J Cereb Blood Flow Metab 27, 597–605. https://doi.org/10.1038/sj.jcbfm.9600358

Mahadevan, A.S., Tooley, U.A., Bertolero, M.A., Mackey, A.P., Bassett, D.S., 2021. Evaluating the sensitivity of functional connectivity measures to motion artifact in resting-state fMRI data. NeuroImage 241, 118408. https://doi.org/10.1016/j.neuroimage.2021.118408

Mertens N, Sunaert S, Van Laere K, Koole M. The Effect of Aging on Brain Glucose Metabolic Connectivity Revealed by [18F]FDG PET-MR and Individual Brain Networks. Front Aging Neurosci. 2022; 13:798410. doi: 10.3389/fnagi.2021.798410

Messaritaki, E., Dimitriadis, S.I., Jones, D.K., 2019. Optimization of graph construction can significantly increase the power of structural brain network studies. NeuroImage 199, 495–511. https://doi.org/10.1016/j.neuroimage.2019.05.052

Mukaka, M.M., 2012. Statistics Corner: A guide to appropriate use of Correlation coefficient in medical research. Scientific Data 24, 69–71.

Nicastro, N., Stripeikyte, G., Assal, F., Garibotto, V., Blanke, O., 2021. Premotor and fronto-striatal mechanisms associated with presence hallucinations in dementia with Lewy bodies. Neuroimage Clin 32, 102791. https://doi.org/10.1016/j.nicl.2021.102791

Osmanlıoğlu, Y., Alappatt, J.A., Parker, D., Verma, R., 2020. Connectomic consistency: a systematic stability analysis of structural and functional connectivity. J. Neural Eng. 17, 045004. https://doi.org/10.1088/1741-2552/ab947b

Perovnik, M., Rus, T., Schindlbeck, K.A., Eidelberg, D., 2023. Functional brain networks in the evaluation of patients with neurodegenerative disorders. Nat Rev Neurol 19, 73–90. https://doi.org/10.1038/s41582-022-00753-3

Ripp, I., Emch, M., Wu, Q., Lizarraga, A., Udale, R., von Bastian, C.C., Koch, K., Yakushev, I., 2022. Adaptive working memory training does not produce transfer effects in cognition and neuroimaging. Transl Psychiatry 12, 512. https://doi.org/10.1038/s41398-022-02272-7

Ripp, I., Wallenwein, L.A., Wu, Q., Emch, M., Koch, K., Cumming, P., Yakushev, I., 2021. Working memory task induced neural activation: A simultaneous PET/fMRI study. NeuroImage 237, 118131. https://doi.org/10.1016/j.neuroimage.2021.118131

Rolls, E.T., Joliot, M., Tzourio-Mazoyer, N., 2015. Implementation of a new parcellation of the orbitofrontal cortex in the automated anatomical labeling atlas. NeuroImage 122, 1–5. https://doi.org/10.1016/j.neuroimage.2015.07.075

Romero-Garcia, R., Whitaker, K.J., Váša, F., Seidlitz, J., Shinn, M., Fonagy, P., Dolan, R.J., Jones, P.B., Goodyer, I.M., Bullmore, E.T., Vértes, P.E., 2018. Structural covariance networks are coupled to expression of genes enriched in supragranular layers of the human cortex. NeuroImage 171, 256–267. https://doi.org/10.1016/j.neuroimage.2017.12.060

Sala, A., Lizarraga, A., Caminiti, S.P., Calhoun, V.D., Eickhoff, S.B., Habeck, C., Jamadar, S.D., Perani, D., Pereira, J.B., Veronese, M., Yakushev, I., 2023. Brain connectomics: time for a molecular imaging perspective? Trends in Cognitive Sciences S136466132200300X. https://doi.org/10.1016/j.tics.2022.11.015

Seghier ML, Price CJ. Interpreting and utilising intersubject variability in brain function. Trends Cogn Sci. 2018 Jun;22(6):517-530. doi: 10.1016/j.tics.2018.03.003

Shehzad, Z., Kelly, A.M.C., Reiss, P.T., Gee, D.G., Gotimer, K., Uddin, L.Q., Lee, S.H., Margulies, D.S., Roy, A.K., Biswal, B.B., Petkova, E., Castellanos, F.X., Milham, M.P., 2009. The Resting Brain: Unconstrained yet Reliable. Cerebral Cortex 19, 2209–2229. https://doi.org/10.1093/cercor/bhn256

Smith, R., Raffelt, D., Tournierc, J.-D., Connelly, A., 2022. Quantitative streamlines tractography: methods and inter-subject normalisation. Aperture Neuro 1–25. https://doi.org/10.52294/apertureneuro.2022.2.neod9565

Smith, S.M., 2002. Fast robust automated brain extraction. Hum. Brain Mapp. 17, 143–155. https://doi.org/10.1002/hbm.10062

Somandepalli, K., Kelly, C., Reiss, P.T., Zuo, X.-N., Craddock, R.C., Yan, C.-G., Petkova, E., Castellanos, F.X., Milham, M.P., Di Martino, A., 2015. Short-term test–retest reliability of resting state fMRI metrics in children with and without attention-deficit/hyperactivity disorder. Developmental Cognitive Neuroscience 15, 83–93. https://doi.org/10.1016/j.dcn.2015.08.003

Stoessl, A.J., 2017. Glucose utilization: still in the synapse. Nat Neurosci 20, 382–384. https://doi.org/10.1038/nn.4513

Tozzi, L., Fleming, S.L., Taylor, Z.D., Raterink, C.D., Williams, L.M., 2020. Test-retest reliability of the human functional connectome over consecutive days: identifying highly reliable portions and assessing the impact of methodological choices. Network Neuroscience 4, 925–945. https://doi.org/10.1162/netn_a_00148

Tsai, S.-Y., 2018. Reproducibility of structural brain connectivity and network metrics using probabilistic diffusion tractography. Sci Rep 8, 11562. https://doi.org/10.1038/s41598-018-29943-0

van den Heuvel, M.P., Sporns, O., 2011. Rich-Club Organization of the Human Connectome. J. Neurosci. 31, 15775–15786. https://doi.org/10.1523/JNEUROSCI.3539-11.2011

Wang, J., He, Y., 2024. Toward individualized connectomes of brain morphology. Trends in Neurosciences 47, 106–119. https://doi.org/10.1016/j.tins.2023.11.011

Welton, T., Kent, D.A., Auer, D.P., Dineen, R.A., 2015. Reproducibility of Graph-Theoretic Brain Network Metrics: A Systematic Review. Brain Connectivity 5, 193–202. https://doi.org/10.1089/brain.2014.0313

Wu, K., Taki, Y., Sato, K., Kinomura, S., Goto, R., Okada, K., Kawashima, R., He, Y., Evans, A.C., Fukuda, H., 2012. Age-related changes in topological organization of structural brain networks in healthy individuals. Hum. Brain Mapp. 33, 552–568. https://doi.org/10.1002/hbm.21232

Yakushev, I., Drzezga, A., Habeck, C., 2017. Metabolic connectivity: methods and applications. Current Opinion in Neurology 30, 677–685. https://doi.org/10.1097/wco.0000000000000494

Yakushev, I., Ripp, I., Wang, M., Savio, A., Schutte, M., Lizarraga, A., Bogdanovic, B., Diehl-Schmid, J., Hedderich, D.M., Grimmer, T., Shi, K., 2022. Mapping covariance in brain FDG uptake to structural connectivity. Eur J Nucl Med Mol Imaging 49, 1288–1297. https://doi.org/10.1007/s00259-021-05590-y

Yuan, J.P., Henje Blom, E., Flynn, T., Chen, Y., Ho, T.C., Connolly, C.G., Dumont Walter, R.A., Yang, T.T., Xu, D., Tymofiyeva, O., 2019. Test–Retest Reliability of Graph Theoretic Metrics in Adolescent Brains. Brain Connectivity 9, 144–154. https://doi.org/10.1089/brain.2018.0580

Zhao, T., Duan, F., Liao, X., Dai, Z., Cao, M., He, Y., Shu, N., 2015. Test-retest reliability of white matter structural brain networks: a multiband diffusion MRI study. Front. Hum. Neurosci. 9. https://doi.org/10.3389/fnhum.2015.00059

Zuo, X.-N., Xu, T., Milham, M.P., 2019. Harnessing reliability for neuroscience research. Nat Hum Behav 3, 768–771. https://doi.org/10.1038/s41562-019-0655-x

Data and Code Availability Statement

Code is available per reasonable request.

Declaration of Interest Statement

The authors declare that they have no conflicts of interest related to this study.