Received: 8 June 2022 | Revised: 24 August 2022 | Accepted: 7 September 2022

DOI: 10.1002/hbm.26091

RESEARCH ARTICLE

WILEY

Spatial variation of changes in test–retest reliability of functional connectivity after global signal regression: The effect of considering hemodynamic delay

Timothy J. Wanger[1,2] | Amy C. Janes[3] | Blaise B. Frederick[1,2]

[1]McLean Imaging Center, McLean Hospital, Belmont, Massachusetts, USA

[2]Department of Psychiatry, Harvard Medical School, Boston, Massachusetts, USA

[3]Neuroimaging Research Branch, National Institute on Drug Abuse (NIDA), Intramural Research Program, National Institutes of Health, Baltimore, Maryland, USA

Correspondence
Blaise B. Frederick, Department of Psychiatry, Harvard Medical School, Boston, MA 02115, USA.
Email: bbfrederick@mclean.harvard.edu

Funding information
National Institute of Mental Health, Grant/Award Number: 1U54MH091657; National Institute on Drug Abuse, Grant/Award Numbers: DA042987, T32DA015036; National Institutes of Health, Grant/Award Number: R21AG070383

Abstract

Global signal regression (GSR) is a controversial analysis method, since its removal of signal has been observed to reduce the reliability of functional connectivity estimates. Here, we used test–retest reliability to characterize potential differences in spatial patterns between conventional, static GSR (sGSR) and a novel dynamic form of GSR (dGSR). In contrast with sGSR, dGSR models the global signal at a time delay to correct for blood arrival time. Thus, dGSR accounts for greater variation in global signal, removes blood-flow-related nuisance signal, and leaves higher quality neuronal signal remaining. We used intraclass correlation coefficients (ICCs) to estimate the reliability of functional connectivity in 462 healthy controls from the Human Connectome Project. We tested across two factors: denoising method used (control, sGSR, and dGSR), and interacquisition interval (between days, or within session while varying phase encoding direction). Reliability was estimated regionally to identify topographic patterns for each condition. sGSR and dGSR provided global reductions in reliability compared with the non-GSR control. Test–retest reliability was highest in the frontoparietal and default mode regions, and lowest in sensorimotor cortex for all conditions. dGSR provides more effective denoising in regions where both strategies greatly reduce reliability. Both GSR methods substantially reduced test–retest reliability, which was most evident in brain regions that had low reliability prior to denoising. These findings suggest that reliability of interregional correlation is likely inflated by the global signal, which is thought to primarily reflect dynamic blood flow.

KEYWORDS
dense connectome, resting state, test–retest reliability, global signal regression

1 | INTRODUCTION

Resting-state neuroimaging is a flexible technique that has been used to explore intrinsic and dynamic properties of the human brain. Resting-state functional connectivity measures the covariation of blood oxygenation-level-dependent (BOLD) signal between separate brain regions, generally in the absence of an explicit task in order to infer neuronal connectivity. It is advantageous as a paradigm because participants can be successfully scanned with minimal instruction or cooperation, making it especially attractive for studying early stages of cognitive development and in neuropsychiatric populations. This technique is used to identify coactivating brain areas at rest (Biswal

et al., 1995; Buckner et al., 2008; Fox et al., 2005), and has potential as a diagnostic tool for neuropsychiatric disorders (Fox & Greicius, 2010; Greicius et al., 2007). There is strong evidence for the reproducibility of the networks commonly observed via functional connectivity from analogous findings in structural diffusion tensor imaging (Damoiseaux & Greicius, 2009; Van Den Heuvel & Pol, 2010) and independent components analysis (ICA; Zuo et al., 2010). However, concerns remain about the reliability and replicability of functional connectivity estimates. In some cases, the recommended best practices for analyzing resting state data result in a decrease in reliability, while other factors increase it (Noble et al., 2019). In this study, we evaluated the effects of: (1) type of global signal regression (GSR), and (2) interacquisition interval, on test–retest reliability of connectivity calculated from resting state acquisitions in the Human Connectome Project (HCP).

Test–retest reliability, which throughout the rest of the article, we refer to as "reliability," reflects how stable a measurement is between two timepoints of data acquisition and is commonly assessed using the intraclass correlation coefficient (ICC; Shrout & Fleiss, 1979). Highly reliable data suggest replicability of findings but does not confirm the validity of the test metric. Instead, reliability is thought to ascribe an upper bound on the validity of a measurement (for review, see Noble et al., 2019). Reliability of resting state data is affected by scanning parameters, acquisition signal-to-noise ratio (SNR), normal variation of connectivity over short and long timescales, and physiological nuisance fluctuations. Acquiring data at longer durations (Birn et al., 2013; Elliott et al., 2019), with faster sampling (Shah et al., 2016), and shorter between-scan intervals (Birn et al., 2013), have all been shown to improve reliability estimates. In contrast, efforts to denoise data through GSR (Braun et al., 2012; Tozzi et al., 2020), cardiac and respiratory regression (Birn et al., 2014; Shirer et al., 2015), and motion correction (Parkes et al., 2018) generally cause a decrease in estimated reliability of resting state data. These decreases are due to the removal of replicable motion artifacts. Indeed, previous work has established that motion artifacts have characteristic profiles around the edges of the brain and ventricles across subjects (Bianciardi et al., 2009), which are likely to be replicated within-subjects. When taken together, the net result is an increase in reliability when acquiring higher quality data, but also a decrease when high amplitude artifacts, which have spatiotemporal consistency across sessions, are removed via denoising methods. These patterns are a reminder that Functional Magnetic Resonance Imaging (fMRI) is an indirect measure of neural activation that is masked by constant noise from the scanner, movement artifacts, and physiological confounds.

Poor reliability of resting state connectivity reflects a broader issue of reproducibility in neuroimaging (Poldrack et al., 2017), and it is critical to systematically study the impact of analysis techniques as studies using resting-state fMRI become more prevalent. In order to assess the reliability of the underlying neural signal, it is necessary to remove artifactual sources of data that may spuriously inflate the estimation of reliability. Static GSR (sGSR) is a common denoising step that consists of computing the average signal across all brain voxels, and subsequently regressing that time course out of each voxel. Early research suggested that the average global signal was comprised of low-frequency noise generated by cardiac, respiratory, and other physiological sources (Lund et al., 2006; Murphy & Fox, 2017). These theories have been partially supported, with respiration and cardiac frequencies (0.1–0.5 and 0.6–1.2 Hz, respectively) contributing significantly to correlation coefficients in arterial, venous, and cerebrospinal fluid voxels. Thus, removing these signals using an sGSR approach promises to enhance the ability to assess correlations between more pure neuronal signals. However, the source of lower frequency components (0.01–0.1 Hz) of the global signal, which contribute roughly 90% to correlation estimates for cortical voxels (Cordes et al., 2001), is not fully understood. Nevertheless, an incomplete understanding of the global signal has not impeded attempts to remove it through linear regression. This initially raised concerns about how sGSR may mathematically introduce potentially spurious anticorrelations (Fox et al., 2009), but ultimately researchers cautiously encouraged its use within the proper contexts (Murphy & Fox, 2017; Murphy et al., 2009). sGSR offers an improvement to reliability of graph metric estimates in resting state data (Andellini et al., 2015) and enhances the specificity of correlations between regions (Weissenbacher et al., 2009), but may reduce the reliability of regional homogeneity estimates (Zuo et al., 2013).

Whatever the ultimate source of the signal may be, a likely explanation for its pervasive nature is that the global signal underlies all fMRI activity because it is carried by blood, which perfuses the entire brain. This is supported by evidence that the global signal strongly correlates with external measurements of blood parameters such as pressure, heart rate, arterial carbon dioxide, and peripheral fingertip oxygenated hemoglobin (Birn et al., 2006; Erdoğan et al., 2016; Power et al., 2017; Tong et al., 2012; Whittaker et al., 2019; Wise et al., 2004). This issue highlights a complicating factor for signal denoising as the global signals associated with blood flow have been shown to have differential time delays throughout the brain (Christen et al., 2015; Frederick et al., 2012; Tong & Frederick, 2010) when cross-correlated with resting state time courses, indicating a range of blood arrival times in various regions. These findings led to the development of dynamic GSR (dGSR), which also removes the global signal through regression, but with the additional step of using an optimal time-delay in every voxel to account for variable blood arrival times. Recent work has demonstrated improvements to the specificity of both positive and negative correlations for posterior cingulate cortex seed-based connectivity when using dGSR, compared with sGSR (Erdoğan et al., 2016). Additionally, dGSR eliminates ~20% more variance associated with systemic signal compared with sGSR. Based on these assumptions, we hypothesized that the use of a rigorous denoising procedure (dGSR) would reduce reliability of functional connectivity estimates compared with sGSR, by removing a larger amount of reliable hemodynamic artifact, and that this relationship would be most evident in regions associated with high blood volume and early blood flow.

2 | METHODS

2.1 | Participants

Our sample was obtained from the Young Adult dataset (1200 subject release) as part of the HCP (Van Essen et al., 2013). We selected a subset of participants chosen to ensure generalization to neurotypical healthy controls based on exclusion criterion from a prior study (Janes et al., 2020). Briefly, participants were excluded if they had been diagnosed with a psychiatric or neurological disorder, a history of past substance use, or current substance use during scanning sessions, as confirmed by alcohol breathalyzer and urine samples. The final sample included 462 healthy control subjects (281 women $M_{age} = 28.66$). The self-reported ethnic descent of the sample was White ($n = 351$), Black or African American ($n = 64$), Asian, Native Hawaiian, or Pacific Islander ($n = 35$), multiracial ($n = 6$), and unknown or not reported ($n = 6$).

2.2 | Data acquisition

Participants underwent a two-day visit at an HCP research site, where each day included a 2-h scanning session for acquiring structural, task-based, and resting-state MRI. On each day there was a resting-state scanning session (REST1 or REST2), where researchers obtained two 15-min blocks of data that were either left-to-right (LR) or right-to-left (RL) phase encoded, one order following the other. For resting state scans, gradient-echo Echoplanar Images (EPI) were acquired with the following parameters: Repetition time (TR) = 720 ms, Echo time (TE) = 33.1 ms, flip angle = 52°, Field of View (FOV) = 208 × 180 mm, matrix = 104 ×90 mm, resolution = 2 mm isotropic voxels, Multiband factor = 8, and echo spacing 0.58 ms. These data were minimally preprocessed by the HCP consortium (Glasser et al., 2013), and denoised using ICA-FIX (FMRIB's ICA-based Xnoisifier) as described in Salimi-Khorshidi et al. (2014) and Griffanti, et al. (2014). This standardized preprocessing pipeline includes realignment, normalization to Montreal Neurological Institute (MNI) space, spatial smoothing, and subjected to automated removal of motion and other artifacts using ICA-FIX. Furthermore, these data were converted into brainordinates, which remaps the data to "grayordinates"—a parcellation of cortical and subcortical gray matter voxels. For the following analyses, we utilized the grayordinate data.

2.3 | Data analysis overview

We compared the reliability of resting state data across two factors: which denoising processing stream was used (control, sGSR, or dGSR), and interacquisition interval. For the purposes of this study, we refer to the 24 h interval contrast as "day" and the 15-min interval as "phase" due to the co-occuring factor of phase encoding direction. The control group used resting state data that were preprocessed according to (Smith et al., 2013), minimally preprocessed according to the HCP (Glasser et al., 2013) and then had artifacts removed using ICA + FIX (Salimi-Khorshidi et al., 2014) without any form of GSR. Both sGSR and dGSR were performed on ICA-FIX-extended data using the "rapidtide" implementation of the RIPTiDe (Regressor Interpolation at Progressive Time Delays) algorithm (Erdoğan et al., 2016). RIPTiDe denoising models the global mean signal as the sum over the brain of a nuisance time course that is thought to travel with flowing blood, arriving at different voxels at different times. The procedure estimates the traveling signal, determines when it arrives at each voxel, then applies a voxel-wise delay to the waveform to generate and remove voxel-specific nuisance regressors. Importantly, the estimation of moving regressor was performed on unparcellated volumetric data that had not undergone ICA-FIX motion denoising. Aso et al. (2017) showed that ICA denoising methods such as ICA-FIX makes lag estimation unreliable in regions where noise has been removed, as the delay is estimated from the noise itself, and accurate delay estimation is needed for regressor refinement. However, once the moving regressor has been found, we have verified that there is no significant difference in results whether rapidtide denoising is run before or after ICA-FIX on grayordinate data; in areas where ICA-FIX has already removed the noise, rapidtide denoising (and accurate lag determination) is not required. The denoised timeseries data were then parcellated into 400 cortical and 19 subcortical regions using the Shaeffer Atlas (Schaefer et al., 2018) and HCP's 1200-subject parcellation. Parcellation averages the timeseries data of all grayordinate voxels included in each region, which varied in size between 70 and 258 timeseries for our mapping.

Parcellated connectomes were generated for each analysis stream by calculating Pearson's correlation coefficients between the signals of each pair of regions. These analyses can be understood in terms of graph theory, where each region is described as a "node," and any relationship between nodes, such as correlation, is described as an "edge." Others have identified the impact of methodological choices on edge-level estimates of reliability (Noble et al., 2019; Tozzi et al., 2020). These edge-level estimates have conventionally been averaged to describe ICC patterns at broader levels of analysis (Tozzi et al., 2020; Yoo et al., 2019). Ultimately, we sought to understand the impact of denoising techniques, and to characterize spatial variability of reliability when applied to resting state data.

2.4 | Standard pipeline

ICA-FIX-extended data were released by the HCP consortium as grayordinate timeseries files with the suffix (rfMRI_REST*X*_*Y*_Atlas_hp2000_clean.dtseries.nii, where *X* is 1 or 2, and *Y* is RL or LR). We used these files as the starting point for our analysis. The data were parcellated (wb_command -cifti-parcellate) and connectomes were created using HCP's Connectome Workbench (wb_command -cifti-correlation). These data were then imported into MATLAB using cifti analysis tools (https://github.com/Washington-University/cifti-matlab) to organize the data for edge-level and region-level analyses. We then calculated ICCs using a freely available software package in

R (psych-ICC; Revelle, 2017). For region-level analyses, we averaged the edge-level ICCs between each specific region with all other regions. Additionally, for analyses comparing REST1 to REST2, we included ICCs from both phase-encoding directions in the average (and similarly for analyses contrasting phase). Contrasts between denoising methods are masked to include only regions that change with a Bonferroni-corrected $p < .05$.

2.5 | sGSR and dGSR

sGSR was derived by averaging the BOLD signal across all voxels of the brain and regressing the signal out of each voxel. dGSR utilizes an optimal time delay that maximizes the cross-correlation with the GSR signal for each voxel. This ensures that a greater proportion of variance associated with the global signal is removed. Furthermore, this process can be iteratively fine-tuned to estimate the artifact that underlies the global signal. In order to do this, multiple passes through the RIPTiDe algorithm are required. In each pass, the global signal is re-estimated after time-shifting the data in each voxel by the derived time delay for the prior pass. This refines the estimated global signal and maximizes its fit to the data. For these analyses, we used three passes. dGSR is explicitly designed to reject motion-based signal components; during the refinement process, voxels with a time delay of less than ±0.25 s are excluded from regressor generation. Since motion effects (or other, external impulsive nuisance signal sources) will have no time delay between voxels, this means that these components will not make it into the final regressor (which is not the case for sGSR). By removing a signal that closely resembles the low-frequency blood signal at an incorrect time delay (i.e., zero seconds in all voxels), sGSR introduces an anticorrelated copy of the regressed signal at the time delay that should have been used. As we showed (Erdoğan et al., 2016), dGSR avoids this issue by removing the global signal at the proper time delay. A more detailed description of dGSR methods is available in the literature (Erdoğan et al., 2016; Frederick et al., 2012).

2.6 | Intraclass correlation coefficient

Broadly speaking, ICC is considered to be a group-level statistic that refers to the temporal, or intraindividual stability of an index measured across multiple occasions (Zuo & Xing, 2014; Zuo et al., 2014). ICC is calculated as the ratio of between-subject variance (numerator) and total variance, which includes between-subject and within-subject variance (denominator). Thus, ICC is affected by both intraindividual differences and interindividual differences within the sample. According to Shrout and Fleiss (1979), reliability can be estimated using different formulae that can impact the generalizability of results. The version we used for these analyses is ICC (2,1), which has been used in prior work to test the reliability of resting state fMRI obtained at different sites (Friedman et al., 2008). ICC (2,1) assesses the absolute agreement between two raters (scanners) and is on a scale from 0 to 1. This measures the extent to which the correlation coefficients of our data match exactly between the instances along the experimental dimension of interest (e.g., day, phase encoding). By using ICC (2,1) instead of ICC (3,1) we are not controlling for all fixed effects and will obtain smaller test–retest estimates. However, our primary interest was the relative increase in reliability that is obtained by using different denoising methods rather than the ground truth of the reliability estimate per se, and ICC (2,1) is considered to be more conservative and generalizable if the results are intended to apply to other scanners (for further discussion, see Chen et al., 2018; Noble et al., 2021). Reliability estimates are traditionally interpreted on a scale of "poor" (<0.4), "fair" (0.4–0.59), "good" (0.6–0.74), and "excellent" (>0.75; Cicchetti and Sparrow 1981).

2.7 | Bland–Altman plots

Bland–Altman analysis is one way to quantitatively compare two methods of measurement that examines the mean difference and degree of agreement (Giavarina, 2015). Region-level ICC estimates from control, sGSR, and dGSR groups are compared using this technique.

3 | RESULTS

3.1 | Edge-level correlations

sGSR and dGSR shifted the distribution of Pearson correlation coefficients leftward, resulting in distributions centered around a lower mean (Figure 1). The mean (M), between-subjects variance (σ^2), and standard error (SE) of these correlations after each denoising method

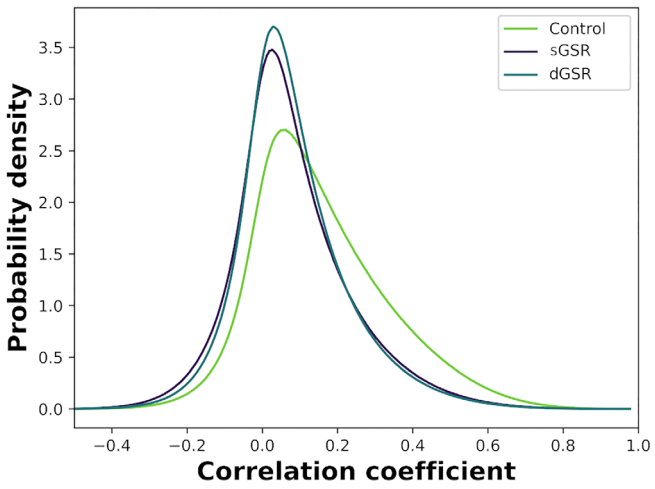

FIGURE 1 Distribution of Pearson correlation coefficients for REST1 versus REST2 comparison. Denoising strategies are depicted in green (control), purple (static global signal regression), and blue (dynamic global signal regression). dGSR, dynamic global signal regression; sGSR, static global signal regression

were: control ($M = 0.1565$, $\sigma^2 = 0.0317$, SE < 0.0001), sGSR ($M = 0.0800$, $\sigma^2 = 0.0245$, SE < 0.0001), and dGSR ($M = 0.0821$, $\sigma^2 = 0.0213$, SE < 0.0001). The mean correlations of the control data were 0.1706 (REST1_LR), 0.1460 (REST1_RL), 0.1444 (REST2_LR), and 0.1649 (REST2_RL). When comparing correlations by condition for the control data, a two-way Analysis of Variance (ANOVA) revealed that there was a significant interaction between the effects of day and phase ($F = 224{,}900$, $p < .0001$).

As shown in Figure 2, when comparing sGSR to control for the "day" contrast, we observed an increase in the percentage of poor (45.3%–49.1%), good (6.3%–10.7%), and excellent edges (0.2%–1.09%) and a decrease in fair edges (48.1%–39.1%). We found similar results in dGSR for poor (45.3%–57.1%), good (6.3%–10.1%), excellent (0.2%–0.98%), and fair edges (48.1%–31.9%).

When comparing sGSR to control for the "phase" contrast, we observed an increase in the percentage of poor (19.8%–27.6%), good (17.1%–18.4%), and excellent edges (0.5%–2.06%) and a decrease in fair edges (62.5%–51.9%). dGSR resulted an increase for poor (19.8%–38.6%) and excellent edges (0.5%–1.79%), and a decrease in good (17.1%–16.0%), and fair edges (62.5%–43.6%).

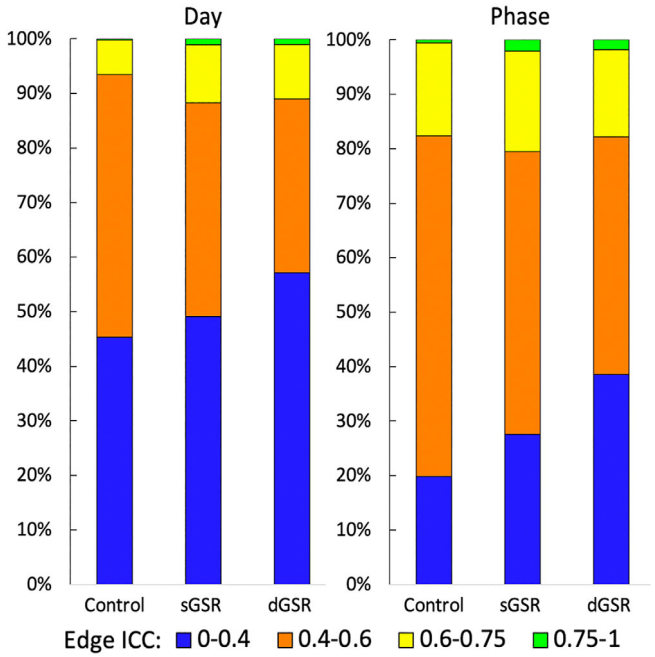

FIGURE 2 Proportion of edges with intraclass correlation coefficients (ICCs) <0.4 (blue), 0.4–0.6 (orange), 0.6–0.75 (yellow), and 0.75–1 (green) as defined by Cicchetti and Sparrow (1981). Three denoising strategies (control, static global signal regression, and dynamic global signal regression) are presented for the day (left) and phase (right) contrasts. dGSR, dynamic global signal regression; sGSR, static global signal regression

3.2 | Region-level ICC estimates

We observed that sGSR on average causes no overall change in regional reliability(Figure 3). In contrast, dGSR causes a very small average decrease in reliability. Interestingly, both techniques appear to preferentially increase the reliability of regions that have the highest initial reliability in the control condition, and conversely, decrease the reliability of regions that have the lowest initial reliability. Unsurprisingly, we observed higher reliability estimates when comparing within-session ICCs (phase contrast) than when comparing across sessions (day contrast; Figure 4). It is worth noting that because distortion correction is not perfect, there may be asymmetries between the within-session scans that would make comparing the two-phase encode directions problematic. However, the fact that the ICC reliability is significantly higher between the within-session scans than between scans of the same phase encode direction in different sessions clearly indicates that this asymmetry is not the dominant source of interscan variability—rather, it is the time interval between scans.

In our region-level analyses, we observed distinctly different patterns in reliability alteration. We observed a strong pattern of decreased reliability in sensorimotor, temporal, occipital cortical regions, and subcortical regions (subcortical data shown in Supplemental Materials S1) in both the day and phase after dGSR processing. In contrast, after sGSR processing, we saw a similar but weaker pattern in the phase condition; in the day condition, we saw only sparse

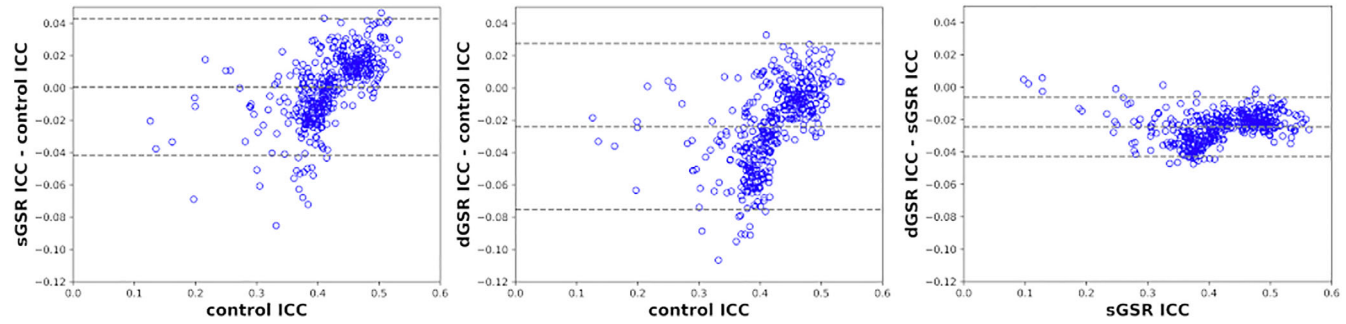

FIGURE 3 Bland–Altman plots comparing the difference in ICC between two methods at the region level. These test the efficacy of denoising techniques at different ranges of reliability (some regions may start at low, or high reliability and be differentially affected by denoising). Horizontal bars indicate the difference in means and 2 SD from the mean difference. ICC, intraclass correlation coefficient; dGSR, dynamic global signal regression; sGSR, static global signal regression

regions of decreased and increased ICC (Figure 5). A direct comparison of the two denoising techniques reveals a pattern of decreased ICC after dGSR processing, with decreases being largest in sensorimotor, temporal, and occipital cortices.

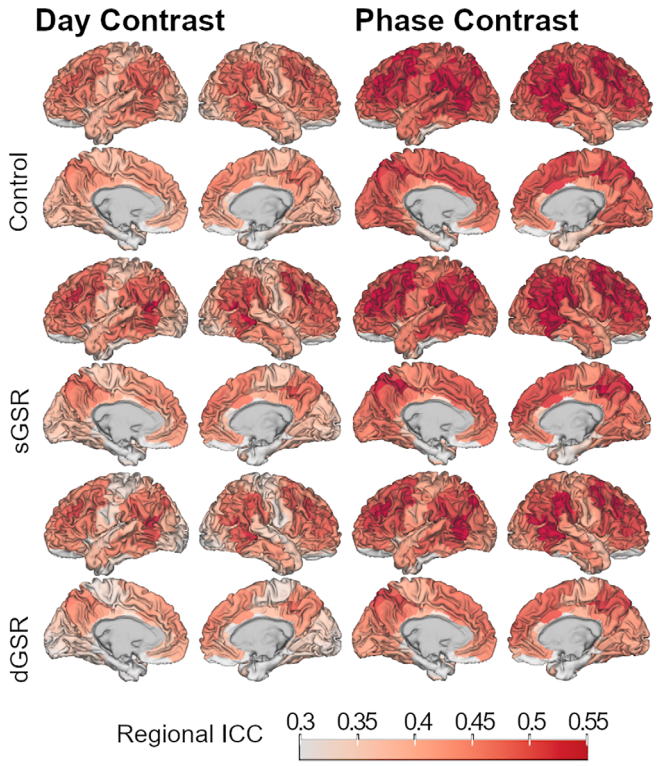

FIGURE 4 Region-level ICC values for the control, static global signal regression, and dynamic global signal regression groups when examining day (REST1 vs. REST2) and phase encoding (left-to-right vs. right-to-left) contrasts. ICC, intraclass correlation coefficient; dGSR, dynamic global signal regression; sGSR, static global signal regression

3.3 | Post hoc correlation analyses

Our region-level analyses suggest that baseline estimates of reliability may vary as a function of spatial location within the brain, and differences between denoising techniques (sGSR and dGSR) may also be spatially based. We used post hoc correlations to test whether these differences are associated with derived parameters of dGSR. First, we tested whether dGSR's maximum Pearson correlation coefficient was related to the difference in reliability between dGSR and sGSR. We observed that regions with a higher maximum correlation value for dGSR (which corresponds to a larger fraction of variance attributed to the moving, blood-borne nuisance signal) were associated with regions had greater relative reductions in ICC ($R = -.39$, $p \leq .001$). Next, we compared dGSR's optimized delay time to the difference in reliability between dGSR and sGSR. We observed a significant correlation ($R = .11$, $p = .026$) showing that regions with lower relative delay times (which corresponds to earlier blood arrival times) were associated with greater relative reductions in ICC (Figure 6).

4 | DISCUSSION

We demonstrated region-specific increases and decreases in reliability after using sGSR and dGSR. Our findings suggest that both increases and decreases in ICC after using sGSR, which have previously been reported (Noble et al., 2019) may be region-specific. These techniques supplement standard preprocessing methods and are effective in removing non-neuronal sources of global signal. In data that have a significant portion of movement artifact removed with ICA-FIX it may be possible to isolate blood-related variance that is not effectively mitigated by standard preprocessing methods. The goal of using denoising tools is to improve validity and reliability estimation of neuronal correlation, even if the strength of the observed correlations is reduced. The stability of the hemodynamic correlation between the

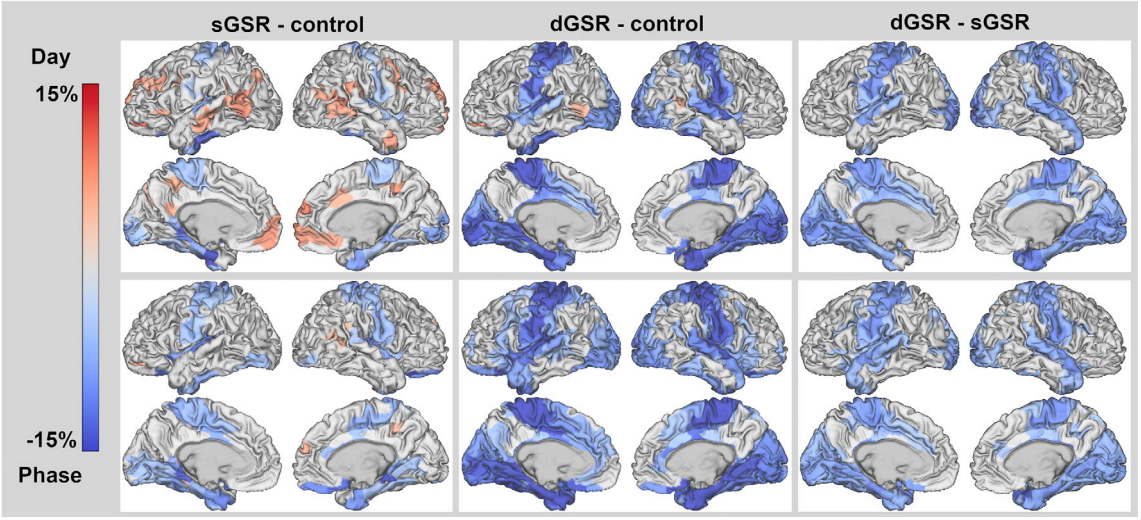

FIGURE 5 Normalized percent change in test–retest reliability for day(top) and phase(bottom) contrasts. From left to right: sGSR–control, dGSR–control, and dGSR–sGSR. The red overlay indicates greater reliability and blue indicates lower reliability. Only changes significant at a Bonferroni-corrected $p < .05$ are colored. dGSR, dynamic global signal regression; sGSR, static global signal regression

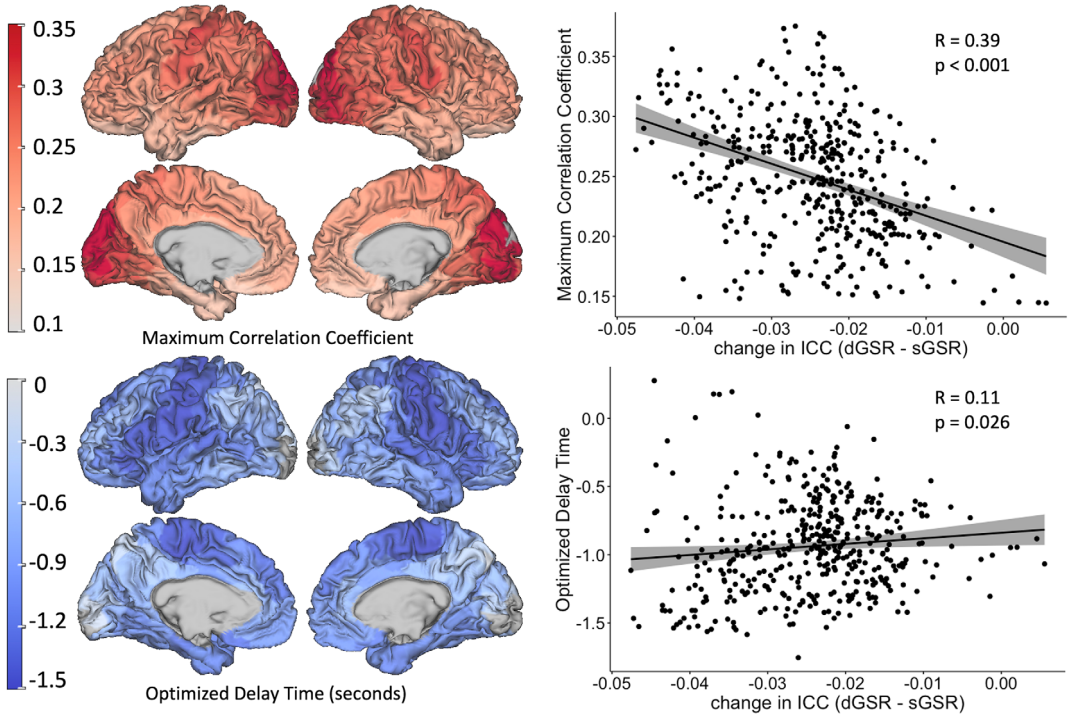

FIGURE 6 The difference in test–retest reliability between denoising methods(dGSR–sGSR) is correlated with two derived parameters of dGSR: Maximum correlation coefficient (top row), and optimized delay time (bottom row). Larger correlation values (red) indicate that more signal is removed from dGSR. Smaller optimized delay times (blue) indicate an earlier offset of the regressed dGSR signal, also corresponding to faster perfusion of blood. dGSR, dynamic global signal regression; sGSR, static global signal regression

regions is based on the stability of the vascular architecture between timepoints. The degree of correlation depends on the vascular structure, not the nuisance signal which is being correlated. If the correlated blood signal between two regions is large compared with the neuronal connectivity between the regions, removing the blood signal will reduce reliability. We expected the largest differences to occur where the most signal is being removed by dGSR. We note that the underlying pattern of low reliability in the sensorimotor cortex and superior temporal gyrus corresponds to the regions that have lower values of optimized delay time from the dGSR analysis (as shown in Figure 6), indicating a potential link between hemodynamic delay and initial low reliability. In contrast, higher reliability was observed in bilateral posterior parietal, middle temporal, superior frontal, and medial frontal cortex. Overall, we conclude that sGSR and dGSR offer likely improvements to the validity of resting state data for estimating nonartifactual, neuronal connectivity, even in cases where the reliability of regions is lowered.

Using sGSR and dGSR had broad impact on the edge strengths, which are used to derive reliability estimates. For researchers who are interested in the "ground truth" of correlational strength between regional activity, the choice to use sGSR or dGSR is very important, and should be carefully considered. We found an overall decrease in median edge strength by using either sGSR or dGSR (Figure 1). These findings are in agreement with the literature, where sGSR has been shown to introduce spurious anticorrelations (Fox et al., 2009; Murphy et al., 2009), whereas dGSR introduces fewer, and smaller, spurious anticorrelations (Erdoğan et al., 2016). Taken together our data demonstrate that the global signal is responsible for inflating the number and strength of positive edge correlations observed in resting state data (Figure 2).

4.1 | Region-level versus edge-level reliability

The impact of sGSR and dGSR on edge proportions are nearly identical to previous work by Tozzi et al. (2020). Reliability has conventionally been examined at the edge level, with few studies attempting to test for effects at the region level. Edge-level statistics may be more generalizable for studies that are examining specific correlations between nodes of a network. However, while individual edges are important for specialized research, it may not be optimal to devise preprocessing protocols that are optimized on these grounds. Examining the reliability of all correlations of a region, as we have employed here, can identify potential underlying regional variations in the ICC metric.

Indeed, we observed symmetrical bilateral patterns of reliability that were apparent when contrasting scan day or phase encoding. This underlying topographical distribution of ICC has not been reported at this level of detail in prior studies. Tozzi et al. (2020) reported similar findings that superior parietal, middle temporal lobes, and dorsolateral prefrontal cortex had the highest proportion of reliable edges. Additionally, a review of the outstanding reliability

literature found that frontoparietal and default mode networks tend to be the most reliable (Noble et al., 2019). Past work has shown that these networks in particular have highly reliable functional architecture across session, scans, and tasks (Finn et al., 2015). This study highlights broadly distributed, coordinated neuronal activity that processes some aspect of ongoing task demands. It is worth noting that the maximum correlation values between the global signal and the region time courses are lower in the frontoparietal regions than the more posterior portions of the cortex, indicating that the contribution of global hemodynamic effects to the signal is lowest in this region, even before correction.

4.2 | Commonalities of sGSR and dGSR

There are small, but significant differences between Pearson correlation distributions that result from using sGSR and dGSR. However, these differences are small when compared with the distribution of the control condition of no regression (Figure 1), and when examining edge proportions (Figure 2). Furthermore, relative decreases in reliability seem to be strongly associated with the baseline level of reliability of each region (Figure 3). When examining the impact of sGSR and dGSR by region, we can draw several conclusions. First, we observed a similar pattern in reliability across the cortex after correction, with a significant reduction in ICC in the sensorimotor cortex, and limited, scattered increases in frontal and parietal regions. Furthermore, these patterns of ICC change resemble the spatial distribution of ICC strength in the control group (Figure 4) in both the day and phase contrast.

4.3 | Differences between sGSR and dGSR

Our data indicate that dGSR yields relative reductions in reliability across the cortex compared with sGSR. These findings are strongest in regions that have low initial reliability (sensorimotor cortex, superior temporal gyrus, and occipital cortex). This pattern is present in both day and phase contrasts. If we were to base our choice of denoising tools solely on the change in ICC and on the reasonable assumption that higher reliability is better, sGSR would be preferable to dGSR. However, a review of the literature indicates that the application of denoising procedures often results in decreases in ICC (Noble et al., 2019). Some have suggested that reductions in reliability due to signal processing of artifacts are not necessarily bad—as they could improve the validity of estimate of the underlying neuronal signal of interest by removing spatially consistent, "reliable" artifacts (Birn et al., 2014; Noble et al., 2021). Furthermore, we know a priori that the regression for dGSR extracts more signal associated with the global mean than sGSR, from all voxels. Therefore, the comparative decrease in reliability that we observe in the dGSR condition is more likely to reflect the correlations of underlying neuronal signal rather than systemic hemodynamic signal. We hesitate to make prescriptive recommendations based on reliability measures, since they can be inflated in the presence of artifacts and have a complex relationship with other experimental sources of variance. Instead, considering the impact of these techniques on several metrics together may be more helpful in determining which method is preferable. Prior studies have demonstrated that dGSR introduces substantially fewer spurious anticorrelations than sGSR (Erdoğan et al., 2016) and suggests that many of the anticorrelations that remain may be neuronal in origin.

The difference in ICC between techniques was related to dGSR's estimated delay times ($r = .11$) and the maximum correlation value for regions ($r = -.39$). To interpret these findings, it is worth noting that regions that have the "lowest" relative delay time would be perfused with blood first, and regions with the highest maximum correlation would have the most signal regressed out. That is to say, that these regions should enjoy the largest benefits from dGSR; our observations demonstrate that these are also the regions with lower ICC. The time lag pattern modeled by dGSR is a phenomenon that is likely to be the result of traveling hemodynamic signals rather than traveling neuronal signals associated with a visual stimulus (Amemiya et al., 2020). Therefore, we conclude that this parameter reflects a confounding blood-flow related physiological artifact. The differences we observed between sGSR and dGSR are small, but may be more stark in populations where blood flow parameters are more variable. Further work is required to determine what effects this confounding factor has on clinical subgroups that have dysregulated blood flow within the brain (i.e., stroke) or altered vascular function due to central nervous system activity and substance use. For instance, Zubieta et al. (2005) showed that tobacco smokers have increased cerebral blood flow in regions associated with memory, craving, and cue salience. It is possible that this type of effect may be quantifiable through dGSR as changes in the maximum correlation (more blood flow signal is flowing to an area) or delay time (a more responsive or faster perfusion of blood).

4.4 | Limitations

One limitation of this study is the reliance on HCP data, whose paradigm includes both 15-min scan intervals and directional phase-encoding as co-occuring variables. Thus, it would require a reduction of power to determine the effect of phase encoding on the data relative to scanning interval. Furthermore, there appears to be a significant order effect in mean correlations observed in the data—which were counterbalanced (RL followed by LR for the first fMRI session, LR followed by RL in the second session) according to the HCP 1200 Subjects release manual. This effect could be a source of variance that confounds our estimates of reliability. We note that this analysis made use of the multiple runs of the standard HCP-YA dataset to assess reliability, in order to both be directly comparable to the work of Tozzi et al., 2020, and to shed light on the effect of denoising on analyses of these data. Future work will assess the effects of these analyses when applied to datasets explicitly collected to evaluate test–retest reliability, such as those from the Consortium for Reliability and Reproducibility (Zuo & Xing, 2014; Zuo et al., 2014). Additionally, other sources of nuisance variance in the MRI signal such as varying

coil temperature (Firbank et al., 2000) may affect SNR, confounding the "ground truth" of reliability estimations. Future studies using test–retest designs should consider the impact that this and other nuisance factors have on outcome metrics such as ICC, and denoising procedures. Moreover, we observed lower between-subjects variance when using denoising methods (compared with control). The numerator and denominator for the calculation of ICC both contain between-subject variance, and as such, these differences may contribute to the changes in ICC we observed. Future work should consider the impact of denoising methods on both between-subjects and within-subjects variance and consider reporting these as outcome metrics to aid interpretation of reliability estimates. Moreover, refining methodology to optimize measurements to ensure they have high interindividual variability may circumvent this potential problem (Zuo et al., 2019). A second limitation is that these data were obtained from the HCP young adult cohort, therefore these findings may not generalize to significantly older or younger populations. Differences in the development of grey and white matter often found in early adolescence may be more represented in this sample than any atrophy that may occur in old age (Tamnes et al., 2013). Furthermore, it is unclear how vascular pathology in older adults may relate to blood-flow-related artifacts or denoising through sGSR or dGSR.

5 | CONCLUSIONS

We present the effects of two global signal denoising methods on the reliability of functional connectivity estimates in resting state fMRI. The global signal for fMRI data is for the most part an artifact arising from non-neuronal physiological processes. Removing this signal should lead to more accurate quantification of local neuronal signals; however, the effect this has on "reliability" as measured by ICC is complex, and there are additional factors that affect what the optimal denoising strategy is. For instance, several studies have indicated that while sGSR is important for removing global artifacts, it may simultaneously introduce spurious anticorrelations (Erdoğan et al., 2016; Fox et al., 2009; Murphy et al., 2009). More work is needed to understand the underlying characteristics of the global signal to fine-tune denoising methods to effectively remove artifactual signal while maintaining the integrity of the neuronal signal. Using dGSR has the benefit of introducing fewer spurious anticorrelations than sGSR (Erdoğan et al., 2016) while producing similar effects on reliability of resting state correlation estimates. Future work should corroborate the underlying bases for the signal removed by dGSR using other methods of estimating blood flow artifacts.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

DATA AVAILABILITY STATEMENT

This project used publicly available data from the Human Connectome Project.

ORCID

Timothy J. Wanger https://orcid.org/0000-0003-1345-0951
Amy C. Janes https://orcid.org/0000-0002-3749-5006
Blaise B. Frederick https://orcid.org/0000-0001-5832-5279

REFERENCES

Amemiya, S., Takao, H., & Abe, O. (2020). Origin of the time lag phenomenon and the global signal in resting-state fMRI. *Frontiers in Neuroscience*, *14*, 596084. https://doi.org/10.3389/fnins.2020.596084

Andellini, M., Cannatà, V., Gazzellini, S., Bernardi, B., & Napolitano, A. (2015). Test-retest reliability of graph metrics of resting state MRI functional brain networks: A review. *Journal of Neuroscience Methods*, *253*, 183–192. https://doi.org/10.1016/j.jneumeth.2015.05.020

Aso, T., Jiang, G., Urayama, S., & Fukuyama, H. (2017). A resilient, non-neuronal source of the spatiotemporal lag structure detected by BOLD signal-based blood flow tracking. *Frontiers in Neuroscience*, *11*, 1–13.

Bianciardi, M., Fukunaga, M., van Gelderen, P., Horovitz, S. G., de Zwart, J. A., Shmueli, K., & Duyn, J. H. (2009). Sources of functional magnetic resonance imaging signal fluctuations in the human brain at rest: A 7 T study. *Magnetic Resonance Imaging*, *27*(8), 1019–1029. https://doi.org/10.1016/j.mri.2009.02.004

Birn, R. M., Cornejo, M. D., Molloy, E. K., Patriat, R., Meier, T. B., Kirk, G. R., Nair, V. A., Meyerand, M. E., & Prabhakaran, V. (2014). The influence of physiological noise correction on test–retest reliability of resting-state functional connectivity. *Brain Connectivity*, *4*(7), 511–522. https://doi.org/10.1089/brain.2014.0284

Birn, R. M., Diamond, J. B., Smith, M. A., & Bandettini, P. A. (2006). Separating respiratory-variation-related fluctuations from neuronal-activity-related fluctuations in fMRI. *NeuroImage*, *31*(4), 1536–1548. https://doi.org/10.1016/j.neuroimage.2006.02.048

Birn, R. M., Molloy, E. K., Patriat, R., Parker, T., Meier, T. B., Kirk, G. R., Nair, V. A., Meyerand, M. E., & Prabhakaran, V. (2013). The effect of scan length on the reliability of resting-state fMRI connectivity estimates. *NeuroImage*, *83*, 550–558. https://doi.org/10.1016/j.neuroimage.2013.05.099

Biswal, B., Zerrin Yetkin, F., Haughton, V. M., & Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magnetic Resonance in Medicine*, *34*(4), 537–541. https://doi.org/10.1002/mrm.1910340409

Braun, U., Plichta, M. M., Esslinger, C., Sauer, C., Haddad, L., Grimm, O., Mier, D., Mohnke, S., Heinz, A., Erk, S., Walter, H., Seiferth, N., Kirsch, P., & Meyer-Lindenberg, A. (2012). Test–retest reliability of resting-state connectivity network characteristics using fMRI and graph theoretical measures. *NeuroImage*, *59*(2), 1404–1412. https://doi.org/10.1016/j.neuroimage.2011.08.044

Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network: Anatomy, function, and relevance to disease. *Annals of the New York Academy of Sciences*, *1124*, 1–38.

Chen, G., Taylor, P. A., Haller, S. P., Kircanski, K., Stoddard, J., Pine, D. S., Leibenluft, E., Brotman, M. A., & Cox, R. W. (2018). Intraclass correlation: Improved modeling approaches and applications for neuroimaging. *Human Brain Mapping*, *39*(3), 1187–1206. https://doi.org/10.1002/hbm.23909

Christen, T., Jahanian, H., Ni, W. W., Qiu, D., Moseley, M. E., & Zaharchuk, G. (2015). Noncontrast mapping of arterial delay and functional connectivity using resting-state functional MRI: A study in Moyamoya patients. *Journal of Magnetic Resonance Imaging*, *41*(2), 424–430. https://doi.org/10.1002/jmri.24558

Cicchetti, D. V, & Sparrow, S. A. (1981). Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behavior. *American Journal of Mental Deficiency*, *86*(2), 127–137.

Cordes, D., Haughton, V. M., Arfanakis, K., Carew, J. D., Turski, P. A., Moritz, C. H., Quigley, M. A., & Meyerand, M. E. (2001). Frequencies contributing to functional connectivity in the cerebral cortex in "resting-state" data. *American Journal of Neuroradiology*, *22*(7), 1326–1333.

Damoiseaux, J. S., & Greicius, M. D. (2009). Greater than the sum of its parts: A review of studies combining structural connectivity and resting-state functional connectivity. *Brain Structure and Function*, *213*(6), 525–533. https://doi.org/10.1007/s00429-009-0208-6

Elliott, M. L., Knodt, A. R., Cooke, M., Kim, M. J., Melzer, T. R., Keenan, R., Ireland, D., Ramrakha, S., Poulton, R., Caspi, A., Moffitt, T. E., & Hariri, A. R. (2019). General functional connectivity: Shared features of resting-state and task fMRI drive reliable and heritable individual differences in functional brain networks. *NeuroImage*, *189*, 516–532. https://doi.org/10.1016/j.neuroimage.2019.01.068

Erdoğan, S. B., Tong, Y., Hocke, L. M., Lindsey, K. P., & deB Frederick, B. (2016). Correcting for blood arrival time in global mean regression enhances functional connectivity analysis of resting state fMRI-BOLD signals. *Frontiers in Human Neuroscience*, *10*, 311. https://doi.org/10.3389/fnhum.2016.00311

Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., Papademetris, X., & Constable, R. T. (2015). Functional connectome fingerprinting: Identifying individuals using patterns of brain connectivity. *Nature Neuroscience*, *18*(11), 1664–1671.

Firbank, M. J., Harrison, R. M., Williams, E. D., & Coulthard, A. (2000). Quality assurance for MRI: Practical experience. *The British Journal of Radiology*, *73*(868), 376–383. https://doi.org/10.1259/bjr.73.868.10844863

Fox, M. D., & Greicius, M. (2010). Clinical applications of resting state functional connectivity. *Frontiers in Systems Neuroscience*, *4*, 19. https://doi.org/10.3389/fnsys.2010.00019

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences*, *102*(27), 9673–9678. https://doi.org/10.1073/pnas.0504136102

Fox, M. D., Zhang, D., Snyder, A. Z., & Raichle, M. E. (2009). The global signal and observed anticorrelated resting state brain networks. *Journal of Neurophysiology*, *101*(6), 3270–3283. https://doi.org/10.1152/jn.90777.2008

Frederick, B. D., Nickerson, L. D., & Tong, Y. (2012). Physiological denoising of BOLD fMRI data using Regressor Interpolation at Progressive Time Delays (RIPTiDe) processing of concurrent fMRI and near-infrared spectroscopy (NIRS). *NeuroImage*, *60*(3), 1913–1923. https://doi.org/10.1016/j.neuroimage.2012.01.140

Friedman, L., Stern, H., Brown, G. G., Mathalon, D. H., Turner, J., Glover, G. H., Gollub, R. L., Lauriello, J., Lim, K. O., Cannon, T., Greve, D. N., Bockholt, H. J., Belger, A., Mueller, B., Doty, M. J., He, J., Wells, W., Smyth, P., Pieper, S., … Potkin, S. G. (2008). Test–retest and between-site reliability in a multicenter fMRI study. *Human Brain Mapping*, *29*(8), 958–972. https://doi.org/10.1002/hbm.20440

Giavarina, D. (2015). Understanding bland altman analysis. *Biochemia Medica*, *25*(2), 141–151. https://doi.org/10.11613/BM.2015.015

Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., Xu, J. L., Jbabdi, S., Webster, M., Polimeni, J. R., Van Essen, D. C., Jenkinson, M., & Wu-Minn HCP Consortium. (2013). The minimal preprocessing pipelines for the human connectome project. *NeuroImage*, *80*, 105–124. https://doi.org/10.1016/j.neuroimage.2013.04.127

Greicius, M. D., Flores, B. H., Menon, V., Glover, G. H., Solvason, H. B., Kenna, H., Reiss, A. L., & Schatzberg, A. F. (2007). Resting-state functional connectivity in major depression: Abnormally increased contributions from subgenual cingulate cortex and thalamus. *Biological Psychiatry*, *62*(5), 429–437. https://doi.org/10.1016/j.biopsych.2006.09.020

Griffanti, L., Salimi-Khorshidi, G., Beckmann, C. F., Auerbach, E. J., Douaud, G., Sexton, C. E., Zsoldos, E., Ebmeier, K. P., Filippini, N., Mackay, C. E., Moeller, S., Xu, J., Yacoub, E., Baselli, G., Ugurbil, K., Miller, K. L., & Smith, S. M. (2014). ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. *NeuroImage*, *95*, 232–247. https://doi.org/10.1016/j.neuroimage.2014.03.034

Janes, A. C., Peechatka, A. L., Frederick, B. B., & Kaiser, R. H. (2020). Dynamic functioning of transient resting-state coactivation networks in the human connectome project. *Human Brain Mapping*, *41*(2), 373–387. https://doi.org/10.1002/hbm.24808

Lund, T. E., Madsen, K. H., Sidaros, K., Luo, W. L., & Nichols, T. E. (2006). Non-white noise in fMRI: Does modelling have an impact? *NeuroImage*, *29*(1), 54–66. https://doi.org/10.1016/j.neuroimage.2005.07.005

Murphy, K., & Fox, M. D. (2017). Towards a consensus regarding global signal regression for resting state functional connectivity MRI. *NeuroImage*, *154*, 169–173. https://doi.org/10.1016/j.neuroimage.2016.11.052

Murphy, K., Birn, R. M., Handwerker, D. A., Jones, T. B., & Bandettini, P. A. (2009). The impact of global signal regression on resting state correlations: Are anti-correlated networks introduced? *NeuroImage*, *44*(3), 893–905. https://doi.org/10.1016/j.neuroimage.2008.09.036

Noble, S., Scheinost, D., & Constable, R. T. (2019). A decade of test-retest reliability of functional connectivity: A systematic review and meta-analysis. *NeuroImage*, *203*, 116157. https://doi.org/10.1016/j.neuroimage.2019.116157

Noble, S., Scheinost, D., & Constable, R. T. (2021). A guide to the measurement and interpretation of fMRI test-retest reliability. *Current Opinion in Behavioral Sciences*, *40*, 27–32. https://doi.org/10.1016/j.cobeha.2020.12.012

Parkes, L., Fulcher, B., Yücel, M., & Fornito, A. (2018). An evaluation of the efficacy, reliability, and sensitivity of motion correction strategies for resting-state functional MRI. *NeuroImage*, *171*, 415–436. https://doi.org/10.1016/j.neuroimage.2017.12.073

Poldrack, R. A., Baker, C. I., Durnez, J., Gorgolewski, K. J., Matthews, P. M., Munafò, M. R., Nichols, T. E., Poline, J.-B., Vul, E., & Yarkoni, T. (2017). Scanning the horizon: Towards transparent and reproducible neuroimaging research. *Nature Reviews Neuroscience*, *18*(2), 115–126.

Power, J. D., Plitt, M., Laumann, T. O., & Martin, A. (2017). Sources and implications of whole-brain fMRI signals in humans. *NeuroImage*, *146*, 609–625. https://doi.org/10.1016/j.neuroimage.2016.09.038

Revelle, W. R. (2017). *psych: Procedures for personality and psychological research*. Northwestern University.

Salimi-Khorshidi, G., Douaud, G., Beckmann, C. F., Glasser, M. F., Griffanti, L., & Smith, S. M. (2014). Automatic denoising of functional MRI data: Combining independent component analysis and hierarchical fusion of classifiers. *NeuroImage*, *90*, 449–468. https://doi.org/10.1016/j.neuroimage.2013.11.046

Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X. N., Holmes, A. J., Eickhoff, S. B., & Yeo, B. T. (2018). Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cerebral Cortex*, *28*(9), 3095–3114. https://doi.org/10.1093/cercor/bhx179

Shah, L. M., Cramer, J. A., Ferguson, M. A., Birn, R. M., & Anderson, J. S. (2016). Reliability and reproducibility of individual differences in functional connectivity acquired during task and resting state. *Brain and behavior*, *6*(5), e00456. https://doi.org/10.1002/brb3.456

Shirer, W. R., Jiang, H., Price, C. M., Ng, B., & Greicius, M. D. (2015). Optimization of rs-fMRI pre-processing for enhanced signal-noise separation, test-retest reliability, and group discrimination. *NeuroImage*, *117*, 67–79. https://doi.org/10.1016/j.neuroimage.2015.05.015

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420–428. https://doi.org/10.1037/0033-2909.86.2.420

Smith, S. M., Beckmann, C. F., Andersson, J., Auerbach, E. J., Bijsterbosch, J., Douaud, G., Duff, E., Feinberg, D. A., Griffanti, L., Harms, M. P., Kelly, M., Laumann, T., Miller, K. L., Moeller, S., Petersen, S., Power, J., Salimi-Khorshidi, G., Snyder, A. Z., Vu, A. T., … WU-Minn HCP Consortium. (2013). Resting-state fMRI in the human connectome project. *NeuroImage*, *80*, 144–168. https://doi.org/10.1016/j.neuroimage.2013.05.039

Tamnes, C. K., Walhovd, K. B., Dale, A. M., Østby, Y., Grydeland, H., Richardson, G., Westlye, L. T., Roddey, J. C., Hagler, D. J., Jr., Due-Tønnessen, P., Holland, D., Fjell, A. M., & Alzheimer's Disease Neuroimaging Initiative. (2013). Brain development and aging: Overlapping and unique patterns of change. *NeuroImage*, *68*, 63–74. https://doi.org/10.1016/j.neuroimage.2012.11.039

Tong, Y., & Frederick, B. D. (2010). Time lag dependent multimodal processing of concurrent fMRI and near-infrared spectroscopy (NIRS) data suggests a global circulatory origin for low-frequency oscillation signals in human brain. *NeuroImage*, *53*(2), 553–564. https://doi.org/10.1016/j.neuroimage.2010.06.049

Tong, Y., Hocke, L. M., Licata, S. C., & Frederick, B. D. (2012). Low-frequency oscillations measured in the periphery with near-infrared spectroscopy are strongly correlated with blood oxygen level-dependent functional magnetic resonance imaging signals. *Journal of Biomedical Optics*, *17*(10), 106004. https://doi.org/10.1117/1.JBO.17.10.106004

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*(3), 925–945. https://doi.org/10.1162/netn_a_00148

Van Den Heuvel, M. P., & Pol, H. E. H. (2010). Exploring the brain network: A review on resting-state fMRI functional connectivity. *European Neuropsychopharmacology*, *20*(8), 519–534. https://doi.org/10.1016/j.euroneuro.2010.03.008

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E., Yacoub, E., Ugurbil, K., & Wu-Minn HCP Consortium. (2013). The WU-Minn human connectome project: An overview. *NeuroImage*, *80*, 62–79. https://doi.org/10.1016/j.neuroimage.2013.05.041

Weissenbacher, A., Kasess, C., Gerstl, F., Lanzenberger, R., Moser, E., & Windischberger, C. (2009). Correlations and anticorrelations in resting-state functional connectivity MRI: A quantitative comparison of preprocessing strategies. *NeuroImage*, *47*(4), 1408–1416. https://doi.org/10.1016/j.neuroimage.2009.05.005

Whittaker, J. R., Driver, I. D., Venzi, M., Bright, M. G., & Murphy, K. (2019). Cerebral autoregulation evidenced by synchronized low frequency oscillations in blood pressure and resting-state fMRI. *Frontiers in Neuroscience*, *13*, 433. https://doi.org/10.3389/fnins.2019.00433

Wise, R. G., Ide, K., Poulin, M. J., & Tracey, I. (2004). Resting fluctuations in arterial carbon dioxide induce significant low frequency variations in BOLD signal. *NeuroImage*, *21*(4), 1652–1664. https://doi.org/10.1016/j.neuroimage.2003.11.025

Yoo, K., Rosenberg, M. D., Noble, S., Scheinost, D., Constable, R. T., & Chun, M. M. (2019). Multivariate approaches improve the reliability and validity of functional connectivity and prediction of individual behaviors. *NeuroImage*, *197*, 212–223. https://doi.org/10.1016/j.neuroimage.2019.04.060

Zubieta, J. K., Heitzeg, M. M., Xu, Y., Koeppe, R. A., Ni, L., Guthrie, S., & Domino, E. F. (2005). Regional cerebral blood flow responses to smoking in tobacco smokers after overnight abstinence. *American Journal of Psychiatry*, *162*(3), 567–577. https://doi.org/10.1176/appi.ajp.162.3.567

Zuo, X. N., & Xing, X. X. (2014). Test-retest reliabilities of resting-state fMRI measurements in human brain functional connectomics: A systems neuroscience perspective. *Neuroscience & Biobehavioral Reviews*, *45*, 100–118. https://doi.org/10.1016/j.neubiorev.2014.05.009

Zuo, X. N., Anderson, J. S., Bellec, P., Birn, R. M., Biswal, B. B., Blautzik, J., Breitner, J. C., Buckner, R. L., Calhoun, V. D., Castellanos, F. X., Chen, A., Chen, B., Chen, J., Chen, X., Colcombe, S. J., Courtney, W., Craddock, R. C., Di Martino, A., Dong, H. M., … Milham, M. P. (2014). An open science resource for establishing reliability and reproducibility in functional connectomics. *Scientific Data*, *1*(1), 1–13. https://doi.org/10.1038/sdata.2014.49

Zuo, X. N., Kelly, C., Adelstein, J. S., Klein, D. F., Castellanos, F. X., & Milham, M. P. (2010). Reliable intrinsic connectivity networks: Test–retest evaluation using ICA and dual regression approach. *NeuroImage*, *49*(3), 2163–2177. https://doi.org/10.1016/j.neuroimage.2009.10.080

Zuo, X. N., Xu, T., & Milham, M. P. (2019). Harnessing reliability for neuroscience research. *Nature Human Behaviour*, *3*(8), 768–771. https://doi.org/10.1038/s41562-019-0655-x

Zuo, X. N., Xu, T., Jiang, L., Yang, Z., Cao, X. Y., He, Y., Zang, Y. F., Castellanos, F. X., & Milham, M. P. (2013). Toward reliable characterization of functional homogeneity in the human brain: Preprocessing, scan duration, imaging resolution and computational space. *NeuroImage*, *65*, 374–386. https://doi.org/10.1016/j.neuroimage.2012.10.017

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Wanger, T. J., Janes, A. C., & Frederick, B. B. (2023). Spatial variation of changes in test–retest reliability of functional connectivity after global signal regression: The effect of considering hemodynamic delay. *Human Brain Mapping*, *44*(2), 668–678. https://doi.org/10.1002/hbm.26091