Published in final edited form as:
Neuroimage. 2013 December ; 83: 550–558. doi:10.1016/j.neuroimage.2013.05.099.

The effect of scan length on the reliability of resting-state fMRI connectivity estimates

Rasmus M. Birn[a,b,c,*], **Erin K. Molloy**[a], **Rémi Patriat**[b], **Taurean Parker**[d], **Timothy B. Meier**[c,e], **Gregory R. Kirk**[f], **Veena A. Nair**[e], **M. Elizabeth Meyerand**[b,c,e], and **Vivek Prabhakaran**[a,c,e]

[a]Department of Psychiatry, University of Wisconsin-Madison, Madison, WI, USA

[b]Department of Medical Physics, University of Wisconsin-Madison, Madison, WI, USA

[c]Neurosciences Training Program, University of Wisconsin-Madison, Madison, WI, USA

[d]Department of Mathematics, University of Rochester, Rochester, NY, USA

[e]Department of Radiology, University of Wisconsin-Madison, Madison, WI, USA

[f]Waisman Laboratory for Brain Imaging and Behavior, University of Wisconsin-Madison, WI, USA

Abstract

There has been an increasing use of functional magnetic resonance imaging (fMRI) by the neuroscience community to examine differences in functional connectivity between normal control groups and populations of interest. Understanding the reliability of these functional connections is essential to the study of neurological development and degenerate neuropathological conditions. To date, most research assessing the reliability with which resting-state functional connectivity characterizes the brain's functional networks has been on scans between 3 and 11 min in length. In our present study, we examine the test–retest reliability and similarity of resting-state functional connectivity for scans ranging in length from 3 to 27 min as well as for time series acquired during the same length of time but excluding half the time points via sampling every second image. Our results show that reliability and similarity can be greatly improved by increasing the scan lengths from 5 min up to 13 min, and that both the increase in the number of volumes as well as the increase in the length of time over which these volumes was acquired drove this increase in reliability. This improvement in reliability due to scan length is much greater for scans acquired during the same session. Gains in intersession reliability began to diminish after 9–12 min, while improvements in intrasession reliability plateaued around 12–16 min. Consequently, new techniques that improve reliability across sessions will be important for the interpretation of longitudinal fMRI studies.

Keywords

Resting-state; Functional connectivity; fMRI; Reliability; Scan duration; Scan length

[*]Corresponding author at: Department of Psychiatry, University of Wisconsin-Madison, 6001 Research Park Blvd., Madison, WI 53719, USA. Fax: +1 608 263 9340. rbirn@wisc.edu (R.M. Birn).

Conflict of interest
None.

Introduction

Resting state functional connectivity MRI (rs-fcMRI) measures functional connections in the brain via the temporal correlation of low-frequency (<0.1 Hz) fluctuations in the MRI signal. These fluctuations are believed to reflect synchronized variations in spontaneous neuronal firing and unconstrained mental activity (e.g., mind wandering) (Biswal et al., 1995; Fox and Raichle, 2007; Mason et al., 2007). The ability to measure the brain's functional connections noninvasively and in the absence of an instructed task has great utility for neuroscience research and clinical applications. Specifically, the measures of functional connectivity are not influenced by differences in task demand or performance making rs-fcMRI scans relatively easy to acquire and particularly suitable for clinical, pediatric, and aging populations (e.g. Dosenbach et al., 2010; Meier et al., 2012). In large part due to these benefits, there has been a dramatic increase in the number of studies collecting resting-state fMRI data in recent years.

One method of analyzing resting-state fMRI data in order to obtain estimates of functional connectivity is to compute the correlation of the mean signal over an ROI with the signal intensity time-course of every other voxel in the brain (Biswal et al., 1995). While previous studies have shown the primary anatomical relationships in resting-state networks (also referred to as intrinsic connectivity networks, or ICNs) to be similar in scans of healthy individuals (Fox et al., 2005), other studies have shown that cognition, emotional state, and resting condition could influence the consistency with which resting-state fMRI measures ICNs (Harrison et al., 2008; McAvoy et al., 2008; Yan et al., 2009). In addition, resting-state fMRI data is influenced by head motion, cardiac and respiratory effects, and signals from white matter (WM)and cerebrospinal fluid (CSF). These signal fluctuations from non-neuronal activity may overestimate the strength of connectivity and, in turn, affect the reliability of estimating ICNs (Birn, 2012; Power et al., 2012; Weissenbacher et al., 2009).

An important practical question, given these various sources of noise, is how much data to acquire. Most current studies using rs-fcMRI acquire 5–7 min of resting-state fMRI data. This duration is based on earlier studies showing that the strength of functional connectivity is stable from this amount of data (Van Dijk et al., 2010). In addition, recent test–retest rs-fcMRI studies have examined the reliability of functional connectivity for scans between 3 and 11 min in length (Braun et al., 2012; Li et al., 2012; Shehzad et al., 2009; Thomason et al., 2011). However, a recent study by Anderson et al. showed that 12–20 min of resting-state fMRI data are required in order to accurately distinguish the functional connectivity of one individual from a group of subjects using an automated machine-learning classifier (Anderson et al., 2011). In this study, we therefore investigate the influence of scan length, not only on the ability to get similar functional connectivity maps across sessions and subjects, but also on the ability to distinguish individual differences in the strength of functional connections within these networks. Assessing how imaging duration affects reliability is essential for the characterization of individual differences in ICNs as well as changes in brain networks during neurological development or neuropathological degeneration.

Methods

Data acquisition and preprocessing

Three sets of functional scans were collected from 25 participants (ages 35 ± 18 years, 10 females) with no history of neurological or psychological disorders. These sets were comprised of three 10-minute functional scans, each with a different resting condition: eyes closed, eyes open, or eyes fixated on a cross, for a total of nine scans per subject. Two of these sets were acquired within the same imaging session while the final set was acquired two to three months later (Fig. 1). The order of the resting conditions was counterbalanced across imaging sessions and subjects, and informed consent was obtained from all participants before every session in accordance with a Wisconsin Institutional Review Board approved protocol.

Data were collected on a 3.0 T MRI scanner (Discovery MR750) with an 8-channel receive-only RF head coil array (General Electric Medical Systems, Waukesha, WI, USA). All functional images had 3.5 mm isotropic voxels and were acquired sagittally with an echo planar imaging (EPI) sequence: TR = 2.6 s, TE = 25 ms, flip angle = 60°, FOV = 224 mm × 224 mm, matrix size = 64 × 64, slice thickness = 3.5 mm, and number of slices = 40. T1-weighted structural images were collected prior to functional scans with an axial MPRAGE sequence: TR = 8.13 ms, TE = 3.18 ms, TI: 450 ms, flip angle = 12°, FOV = 256 mm × 256 mm, matrix size = 256 × 256, slice thickness = 1 mm, and number of slices = 156.

Functional data were corrected for motion (using AFNI's rigid-body volume registration) and physiological noise from cardiac and respiratory pulsations (RETROICOR) before being aligned to their respective anatomical image (Cox, 1996; Glover et al., 2000; Saad et al., 2009). Automated segmentation (FSL's FAST) of the T1-weighted structural image was used to define masks of the WM and CSF (Smith et al., 2004; Woolrich et al., 2009; Zhang et al., 2001). The two signal intensity time-courses resulting from averaging the fMRI data within the eroded WM and CSF masks as well as their first derivatives (computed by backward difference) were taken as signals of no-interest (i.e., spurious fluctuations unlikely to be of neuronal origin) and removed from the functional data along with the six rigid-body motion registration parameters (Friston et al., 1996; Weissenbacher et al., 2009). The functional images were then transformed to Talairach Atlas space (Talairach and Tournoux, 1988), temporally band-pass filtered between 0.01 Hz and 0.1 Hz (Cordes et al., 2001; Nir et al., 2008), and spatially smoothed with a 3-dimensional Gaussian kernel (FWHM = 4 mm).

Functional connectivity

Functional connectivity was computed between 18 different regions in the brain (Supplementary Table 1), which were defined within the auditory, default mode, dorsal attention, motor, and visual networks (Van Dijk et al., 2010). In each of these regions, signals of interest were extracted from each of the 10-minute functional scans (e.g., eyes closed, fixated, eyes opened) by averaging the preprocessed signal over spherical ROIs (radius = 4 mm). These time courses were normalized after their means were removed and combined into new time series varying in length. Specifically, equal durations of eyes closed, eyes open, and eyes fixated resting runs collected within the same session were

concatenated in the same order as their acquisition to create time series with lengths of 3, 6, 9, 12, 15, 18, 21, 24, and 27 min for each ROI respectively (Fig. 2). For each of these lengths, a connectivity matrix was generated by computing the correlation coefficients (via linear regression using AFNI's 3d Deconvolve function) of each time series with each of the other 17 signals of interest. In this regression, the baseline and linear trends were modeled separately for each of the concatenated resting conditions. That is, the design matrix included the seed time series as well as six additional regressors that model the baseline and linear trend in each of the concatenated runs. The Pearson's correlation in this case is determined from the partial-R^2 which reflects the unique variance explained by the seed time series beyond what is modeled by the baseline regressors. The resulting Pearson's correlation coefficients were converted into a more normal distribution via the Fisher's Z transform. This process was repeated for each of a subject's three sets of scans, and the unique 153 connectivity values from time series with the same scan length were combined into a single multi-scan connectivity matrix (size = 153 × 3) for each subject. Columns one and two (sets of scans acquired during the same imaging session) were used to assess intrasession reliability and columns one and three (sets of scans acquired during different imaging sessions)were used to assess considered intersession reliability (Fig. 2).

In order to evaluate whether the observed improvements in reliability for longer scan lengths were solely due to the greater number of acquired images, connectivity matrices were also generated from time series with lengths of 6, 12, and 24 min by including only half of the number of time points. Specifically, only every second time point (i.e., sampling every 5.2 s or twice the TR) in the time series was included in the regression analysis. These matrices are denoted 3/6, 6/12, and 12/24, where the numerator indicates the amount of time (in minutes with the original TR of 2.6 s) necessary to acquire the number of time points included in the regression, and the total imaging duration (in minutes) from which those time points were taken is given by the denominator. In other words, 3/6 denotes a matrix where every other data point was sampled (one volume every 5.2 s) for a total duration of 6 min. Thus, a 3/6 data set has the same number of time points as a fully sampled 3 minute run (3/3), but spread over a 6 minute duration. This sub-sampling analysis was performed on data that was not temporally filtered, in order to avoid potential influences of band-pass filtering (0.01–0.1 Hz) on the effects of sub-sampling the data.

In light of recent publications demonstrating subtle head motion's effect on resting-state functional connectivity (Power et al., 2012; Satterthwaite et al., 2012; Van Dijk et al., 2012), a repeated measures ANOVA on the mean volume-to-volume displacement of each time series was computed and showed no significant differences across the nine scan lengths ($p = 0.867$). Similarly, a two-way repeated measures ANOVA (categories: scan length and number of time points) on the volume-to-volume displacement of time series from full and "halved" time series of 6, 12, and 24 min also showed no significant differences in motion related to either the scan length ($p = 0.998$) or the number of time points in the time series ($p = 0.763$). The volume-to-volume displacement was estimated from the 6 rigid body motion registration parameters. Specifically, the motion (in mm) between consecutive time points i and j was calculated as follows (Jones et al., 2010; Kennedy and Courchesne, 2008; Meier et al., 2012):

$$\sqrt{(x_j-x_i)^2+(y_j-y_i)^2+(z_j-z_i)^2+(\alpha_j-\alpha_i)^2+(\beta_j-\beta_i)^2+(\gamma_j-\gamma_i)^2}.$$

Translations (x, y, z in mm) and rotations (α, β, γ in degrees) are combined in this manner since one degree of rotation at the center of the head is approximately 1 mm of movement at the surface of the head.

Reliability of functional connections

The intraclass correlation coefficient (ICC) measures the reliability with which fMRI estimates functional connections in an individual's ICN (Shehzad et al., 2009). For each of the nine scan lengths, the 153 × 1 connectivity matrices from the 25 subjects (see the Functional connectivity section) were rearranged into intrasession and intersession n × k matrices for each connection (i.e., each seed pair), where n is the number of subjects (25) and k is the number of fMRI data sets (2). In other words, both intrasession and intersession ICCs for single ratings with random effects were computed for each connection (matrix) separately using the following equation, where MSb is the between subject mean squares and MSw is the within subject mean squares (Shrout and Fleiss, 1979):

$$\mathrm{ICC}=\frac{\mathrm{MSb}-\mathrm{MSw}}{\mathrm{MSb}+(k-1)\mathrm{MSw}}.$$

When the between subject variance is much greater than the within subject variance, the ICC is close to 1 and connection strength as measured by fMRI scans would be considered reliable.

Differences between functional connections

The root-mean-square deviation (RMSD) measures the mean difference between connectivity values from different fMRI scans of the same individual (Anderson et al., 2011). For each of the nine scan lengths, the intrasession and intersession RMSDs were calculated for each subject using the following equation, where A is a subject multi-scan connectivity matrix, j and k are different fMRI data sets (columns), and m is the total number of connections (rows) in each matrix:

$$\mathrm{RMSD}=\sqrt{\frac{\sum_{i=1}^{m}(A_{ij}-A_{ik})^2}{m}}.$$

Similarity of "connected" functional connections

The Dice coefficient (DC) measures the similarity between sets ranging from 0 to 1, where 0 indicates that the sets are disjoint and 1 indicates that the sets are identical (Dice, 1945; Sørensen, 1948). In functional MRI, the DC has been used to assess the commonality of connectivity values from different scans of the same individual that survive a given threshold, z_t (Craddock et al., 2012). Define the function, which transforms an element x into zero or one, as follows:

$$f(x)=\begin{cases} 1, |x| > z_t \\ 0, |x| \leq z_t \end{cases} .$$

Then the intrasession and intersession DCs can be calculated for each subject using the following equation, where A is a subject multi-scan connectivity matrix, j and k are different fMRI data sets (columns), and m is the total number of connections (rows) in each matrix:

$$DC=\frac{2 \times \sum_{i=1}^{m} f\ (A_{ij})\ f(A_{ik})}{\sum_{i=1}^{m} f\ (A_{ij}) + \sum_{i=1}^{m} f(A_{ik})}.$$

In our study, we used two different thresholds in our calculation of the Dice coefficient: (1) $z_t = 0.3$, which is independent of the number of time points, and (2) z_t such that $p = 0.05$ (Bonferroni corrected for multiple comparisons), which is dependent on the number of time points (NT) in the time series from which the connectivity matrices were generated.

Statistical analysis

The above statistics were calculated for the each of the different scan lengths. The significance of imaging duration on each measure was determined by one-way ANOVAs (variable: scan length; categories: 3, 6, 9, 12, 15, 18, 21, 24, and 27 min). Note that for analyses on ICCs, the targets are connections (i.e., for each run length, there exists an ICC for each of the 153 connections), whereas the targets are subjects for RMSDs and DCs (i.e., for each scan length, there exists an RMSD and a DC for each of the 25 subjects). Therefore, repeated measures ANOVAs were used in the analysis of differences in scan length for RMSD and DC. To examine whether decreasing the number of time points over the same amount of scanning time also decreases reliability, two-way ANOVAs (variables: scan length, number of time points; categories: 3, 6, 12, and 24 min) were performed. These statistical tests were performed including data from all of the 153 connections. In addition, these analyses were performed on ICCs and RMSDs for a subset of network connections whose mean Fisher's Z was positive and significant ($p < 0.05$, Bonferroni-corrected) across all scans with scan lengths of 27 min. This subset was selected since both ICCs and RMSDs are affected by the inclusion of null connections; if two seeds are not "connected," their correlation is expected to be zero with variance between scans coming from noise. In theory, the between and within subject variance of null connections should be the same. Finally, as we observed that test–retest metrics improved with the length of the resting-state fMRI scan but that this improvement diminished over time, additional analyses were performed to assess the beginning of the plateaus, or diminishing returns (Supplementary methods).

Results

Fig. 3 shows the effect of scan length on the reliability and similarity metrics. Fig. 4 compares the effects of both scan length and the number of volumes on the reliability and similarity metrics.

Reliability of functional connections

As shown in Fig. 3A, intrasession reliability is much greater than intersession reliability, both of which significantly increase with scan length when all 153 connections are included in the one-way ANOVA (Table 1). When only significant within network connections are considered, the intrasession reliability still increases significantly with longer scan lengths, while the intersession reliability shows smaller improvements (Fig. 3B). This improvement in intrasession ICCs from increased scan length begins to plateau around 13 min, while gains in the intersession ICCs slow around 9 min (Supplementary Figs. 1 and 2). Specifically, the intrasession ICCs continue to increase for scans beyond 6 min in length, with a 20% gain in ICC for scans of 12 min compared to 6 min. The mean ICC is greater when only significant network connections are included, as expected.

As shown in Fig. 4A, a two-way ANOVA (scan duration and number of volumes; Table 2) on all 153 connections indicates that both the scan duration as well as number of volumes have significant effects on intrasession reliability. When only significant within network connections are considered, only scan duration significantly affects intrasession reliability (Fig. 4B). No significant changes in intersession reliability were associated either scan length or the number of volumes in this ANOVA.

Differences between functional connections

As shown in Figs. 2C and D, both intrasession and intersession RMSDs calculated from all 153 connections, as well as those calculated from only significant network connections, significantly decreased with scan length (Table 1). These improvements in both intrasession and intersession RMSDs from significant network connections began to diminish around 8 and 7 min, respectively (Supplementary Figs. 1 and 2). In congruence with results from examining ICCs, RMSDs are better (lower) for intrasession scans. However, when only significant network connections were considered, the intrasession and intersession RMSDs were worse (higher). Differences in RMSD were significantly associated with both the number of volumes and the scan duration in the two-way repeated measures ANOVA (Table 2, Figs. 4C and D).

Similarity of functional connectivity networks

We first examined the DCs with a constant threshold, $|z| > 0.3$, for each scan length (Fig. 3E). Despite the number of connections with $|z| > 0.3$ decreasing slightly with scan length (Fig. 5A), the intrasession and intersession DCs still significantly increased (Table 1). Thus, the intersection of strongly connected brain regions increases with scan length. Improvements began to slow at 12 and 9 min for intrasession and intersession DCs, respectively (Supplementary Figs. 1 and 2). In addition, the number of network connections with $|z| > 0.3$ as well as anti-correlated network connections with $|z| > 0.3$ were relatively stable, which is consistent the stability of connection strength as well as the identification of primary anatomical relationships in resting-state networks after only 5 min of scanning (Fox et al., 2005; Van Dijk et al., 2010).

As shown in Fig. 3F, DCs calculated from connections with $p < 0.05$ (Bonferroni-corrected) significantly improved with scan length (Table 1), and gains in intrasession and intersession

DCs began to plateau around 16 and 12 min, respectively (Supplementary Figs. 1 and 2). As illustrated in Fig. 5B, the number of significant connections increases with scan length. This is as expected, since increasing scan length increases the number of volumes and hence degrees of freedom included in the calculations of significance. Finally, in the two-way repeated measures ANOVA, scan length had a significant effect on intrasession DCs computed with both a constant threshold and a threshold based on significance (Table 2). The number of volume effect on DCs trends toward significance.

Discussion

Many current resting-state fMRI studies acquire approximately 5–7 min of data to obtain estimates of connectivity. Prior studies have shown that this duration of acquisition results in stable estimates of ICNs (Fox et al., 2005; Van Dijk et al., 2010). A study by Braun et al. found slightly lower, but not significantly lower, reliability of estimating functional connections for scan durations of 200 s and 250 s compared to 300 s (Braun et al., 2012). However, this study did not investigate the reliability of runs longer than 5 min. Our study shows that the reliability of measuring individual differences in the strength of connectivity continues to increase for durations longer than 6 min. In fact, a 12-minute scan resulted in 20% greater intrasession ICC compared to a 6-minute scan for the subset of network connections ($p < 0.05$, Bonferroni-corrected). Similarly, the Dice coefficient (DC) of 12-minute scans increased by 36% and 25% for thresholds of $p = 0.05$ (Bonferroni-corrected) and $|z_t| = 0.3$, respectively. These findings suggest that increasing a 5-minute scan to 12–16 min can improve reliability, and is consistent with work by Anderson et al. (2011). Specifically, we found improvements in test–retest metrics from increasing scan length to 12–16 min and 8–12 min for intrasession and intersession data, respectively. Other measures, such as improved session-to-session alignment and noise correction techniques, are necessary to improve intersession reliability, which is critical for longitudinal studies.

Functional connectivity estimates computed by considering only every other data point of a 6, 12, or 24-minute scan length were less reliable than the full 6, 12, 24-minute acquisitions, but were more reliable than full 3, 6, and 12-minute scan lengths, even though the number of time points was the same. Thus, the improvement in reliability for longer scan lengths was not solely due to the greater number of acquired volumes, but was at least partially dependent on the total duration of the acquisition. A potential explanation for this finding is that the functional connectivity within these networks changes on a very slow time scale. Acquiring only 6 min of resting-state data may provide only a snapshot of these slow changes, which changes for another 6-minute run acquired later in the same session. In contrast, functional connectivity computed from a 12-minute duration acquisition averages over enough slow changes to provide a more stable estimate of the connectivity strength. Of course, these slow changes could be either neuronal or artifactual in nature. This sub-sampling analysis looked at the relative influence of the number of acquired time points, sampling rate, and scan duration on the reliability and similarity of functional connectivity, keeping all other factors the same. However, this does not necessarily mean that the same results would be obtained if the data were actually acquired at a longer (or shorter) TR. Acquiring the data at a different TR would influence the signal due to a variety of additional factors, such as T1 recovery and in-flow effects. The extent to which more rapid imaging

pulse sequences (shorter TRs) improve not only the sensitivity of detecting connections but also the test–retest reliability is an interesting and important topic for future study.

A recent study by Kalcher et al. (2012) looking at the data from the 1000 Functional Connectomes Project (http://www.nitrc.org/projects/fcon_1000/) found a correlation between the median number of components obtained from an independent component analysis of resting-state data and the duration of the scan. It is possible that these results reflect the greater number of functional networks that are reliably detected for longer duration scans, which would be consistent with our findings.

The ICCs improved when only significant network connections were considered. In contrast, the RMSD was lower (better) when all connections were considered (Fig. 4). This could result from a greater variability in the connection strength of within-network connections across sessions compared to null connections (areas that are not supposed to be connected). The greater variability of connections currently considered within the ICNs of healthy adults compared to null connections suggests that the network connection strength is influenced by something other than noise (e.g., cognitive state), which would vary the strength of an individual's connection from session to session. In addition, session-to-session stability in subject connectivity matrices (measured here by the RMSD) do not benefit as greatly from increasing scan length as test–retest reliability. Specifically, improvements in RMSDs plateau around 7–8 min, which is more consistent with a study by Van Dijk et al. (2010). showing connectivity strength to become stable around 6 min.

A potential limitation of our study is that the scan lengths were formed from three fMRI scans with different resting-states (e.g., eyes closed, eyes opened, eyes fixated). The connectivity from these concatenated runs would effectively represent an average connectivity across the different resting conditions. While differences in connectivity strength and reliability across resting conditions may be statistically significant for some networks, the differences are relatively small (Patriat et al., 2013). In addition, since equal amounts of resting-state were included in each scan length, alterations in connectivity due to resting condition should be the same across all run lengths and not change the directionality of our findings. Furthermore, repeating this analysis separately for each resting condition resulted in relationships of reliability and similarity with scan length that were similar across different resting conditions (Supplementary Table 2, Supplementary Figs. 3 and 4). Since it was necessary to compose each scan length from equal amounts of the three resting conditions, the scan lengths did not occur consecutively in time. In fact, a scan length of 3-minutes was composed of three 1-minute runs that occurred over a time span of at least 21 min. However, the reliability and similarity of various contiguous data durations showed a similar behavior (Supplementary Figs. 3, 4).

A second limitation to our study was that in order to preserve the number of time points for our comparisons, we did not censor out time points possibly corrupted by motion. Recent studies showing the implications of even subtle head movement on ICNs suggest that head motion could impact reliability. However, as there were no significant differences in the volume-to-volume motion across the different scan lengths, there is no reason to believe head motion would have biased the reliability in favor of particular scan lengths.

Our study focused on the test–retest reliability of strength of connectivity measured by seed-based connectivity (specifically the Z-transformed correlation coefficient between regions of interest). Other functional connectivity metrics may have different levels of reliability. For example, Braun et al. (2012) found that second order graph theory metrics (e.g. small-worldness, hierarchy, assortativity) were in generalmore reliable. A recent study by Zuo et al. (2013) showed that the ICCs of a regional homogeneity (ReHo) metric in the posterior cingulate and precuneus brain regions increased monotonically with scan durations of 1 min to 10 min, but that an inter-session ICC of 0.5 could be achieved with only 4 min of data. This reliability varied across the brain, and it is therefore possible the ReHo measures in some regions of the brain benefit further from longer duration acquisitions.

Recent test–retest fMRI studies have all found intersession ICCs to be consistently lower than those from intrasession scans, and our study is no exception. Furthermore, we find that intersession scans do not benefit as greatly from increased scan length. However, there are many confounding variables that affect the reliability of intersession data. Changes in cognitive and emotional states may activate different brain regions increasing the variability in functional connectivity networks within the same subject. In addition, the position of the brain during the scan and differences in physiological noise (e.g., changes in blood pressure or caffeine intake) could affect the reliability. The development of processing techniques to increase intersession subject network consistency and similarity would be valuable for longitudinal fMRI studies.

Conclusion

Our results show that while it is standard to use 5–7 min of resting-state data, the test–retest reliability and across-session similarity of functional connectivity estimates can be greatly improved by increasing the imaging duration to 9–13 min or longer. In addition, we found that the improved reliability for longer scan lengths was not solely due to the increased number of time points, but also due to the total duration of the scan. This suggests that resting-state functional connectivity estimates may be modulated by slow frequency dynamics, with cycles on the order of several minutes. Further research is needed to determine the source of these dynamics, and new techniques meant to improve intersession reliability are necessary to improve the interpretation of resting-state functional connectivity changes from longitudinal fMRI studies.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neuroimage.2013.05.099.

Acknowledgments

We thank Alexander J. Shackman for the valuable discussions about the analyses. We also thank Bharat B. Biswal for his assistance in the study design. This research was supported by NIH grant RC1MH090912 and the Health Emotions Research Institute.

References

Anderson JS, Ferguson MA, Lopez-Larson M, Yurgelun-Todd D. Reproducibility of single-subject functional connectivity measurements. AJNR Am J Neuroradiol. 2011; 32:548–555. [PubMed: 21273356]

Birn RM. The role of physiological noise in resting-state functional connectivity. Neuroimage. 2012; 62:864–870. [PubMed: 22245341]

Biswal B, Yetkin FZ, Haughton VM, Hyde JS. Functional connectivity in themotor cortex of resting human brain using echo-planar MRI. Magn Reson Med. 1995; 34:537–541. [PubMed: 8524021]

Braun U, Plichta MM, 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. Test–retest reliability of resting-state connectivity network characteristics using fMRI and graph theoretical measures. Neuroimage. 2012; 59:1404–1412. [PubMed: 21888983]

Cordes D, Haughton VM, Arfanakis K, Carew JD, Turski PA, Moritz CH, Quigley MA, Meyerand ME. Frequencies contributing to functional connectivity in the cerebral cortex in "resting-state" data. AJNR Am J Neuroradiol. 2001; 22:1326–1333. [PubMed: 11498421]

Cox RW. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res. 1996; 29:162–173. [PubMed: 8812068]

Craddock RC, James GA, Holtzheimer PE III, Hu XP, Mayberg HS. A whole brain fMRI atlas generated via spatially constrained spectral clustering. Hum Brain Mapp. 2012; 33:1914–1928. [PubMed: 21769991]

Dice LR. Measures of the amount of ecologic association between species. Ecology. 1945; 26:297–302.

Dosenbach NU, Nardos B, Cohen AL, Fair DA, Power JD, Church JA, Nelson SM, Wig GS, Vogel AC, Lessov-Schlaggar CN, Barnes KA, Dubis JW, Feczko E, Coalson RS, Pruett JR Jr, Barch DM, Petersen SE, Schlaggar BL. Prediction of individual brain maturity using fMRI. Science. 2010; 329:1358–1361. [PubMed: 20829489]

Fox MD, Raichle ME. Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. Nat Rev Neurosci. 2007; 8:700–711. [PubMed: 17704812]

Fox MD, Snyder AZ, Vincent JL, Corbetta M, Van Essen DC, Raichle ME. The human brain is intrinsically organized into dynamic, anti correlated functional networks. Proc Natl Acad Sci U S A. 2005; 102:9673–9678. [PubMed: 15976020]

Friston KJ, Williams S, Howard R, Frackowiak RS, Turner R. Movement-related effects in fMRI time-series. Magn Reson Med. 1996; 35:346–355. [PubMed: 8699946]

Glover GH, Li TQ, Ress D. Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. Magn Reson Med. 2000; 44:162–167. [PubMed: 10893535]

Harrison BJ, Pujol J, Ortiz H, Fornito A, Pantelis C, Yucel M. Modulation of brain resting-state networks by sad mood induction. PLoS One. 2008; 3:e1794. [PubMed: 18350136]

Jones TB, Bandettini PA, Kenworthy L, Case LK, Milleville SC, Martin A, Birn RM. Sources of group differences in functional connectivity: an investigation applied to autism spectrum disorder. Neuroimage. 2010; 49:401–414. [PubMed: 19646533]

Kalcher K, Huf W, Boubela RN, Filzmoser P, Pezawas L, Biswal B, Kasper S, Moser E, Windischberger C. Fully exploratory network independent component analysis of the 1000 functional connectomes database. Front Hum Neurosci. 2012; 6:301. [PubMed: 23133413]

Kennedy DP, Courchesne E. The intrinsic functional organization of the brain is altered in autism. Neuroimage. 2008; 39:1877–1885. [PubMed: 18083565]

Li Z, Kadivar A, Pluta J, Dunlop J, Wang Z. Test–retest stability analysis of resting brain activity revealed by blood oxygen level-dependent functional MRI. J Magn Reson Imaging. 2012; 36:344–354. [PubMed: 22535702]

Mason MF, Norton MI, Van Horn JD, Wegner DM, Grafton ST, Macrae CN. Wandering minds: the default network and stimulus-independent thought. Science. 2007; 315:393–395. [PubMed: 17234951]

McAvoy M, Larson-Prior L, Nolan TS, Vaishnavi SN, Raichle ME, d'Avossa G. Resting states affect spontaneous BOLD oscillations in sensory and paralimbic cortex. J Neurophysiol. 2008; 100:922–931. [PubMed: 18509068]

Meier TB, Desphande AS, Vergun S, Nair VA, Song J, Biswal BB, Meyerand ME, Birn RM, Prabhakaran V. Support vectormachine classification and characterization of age-related reorganization of functional brain networks. Neuroimage. 2012; 60:601–613. [PubMed: 22227886]

Nir Y, Mukamel R, Dinstein I, Privman E, Harel M, Fisch L, Gelbard-Sagiv H, Kipervasser S, Andelman F, Neufeld MY, Kramer U, Arieli A, Fried I, Malach R. Interhemispheric correlations of slow spontaneous neuronal fluctuations revealed in human sensory cortex. Nat Neurosci. 2008; 11:1100–1108. [PubMed: 19160509]

Patriat R, Molloy EK, Meier TB, Kirk GR, Nair VA, Meyerand ME, Prabhakaran V, Birn RM. The effect of resting condition on resting-state fMRI reliability and consistency: a comparison between resting with eyes open, closed, and fixated. Neuroimage. 2013; 78:463–473. [PubMed: 23597935]

Power JD, Barnes KA, Snyder AZ, Schlaggar BL, Petersen SE. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. Neuroimage. 2012; 59:2142–2154. [PubMed: 22019881]

Saad ZS, Glen DR, Chen G, Beauchamp MS, Desai R, Cox RW. A new method for improving functional-to-structural MRI alignment using local Pearson correlation. Neuroimage. 2009; 44:839–848. [PubMed: 18976717]

Satterthwaite TD, Wolf DH, Loughead J, Ruparel K, Elliott MA, Hakonarson H, Gur RC, Gur RE. Impact of in-scanner head motion on multiple measures of functional connectivity: relevance for studies of neurodevelopment in youth. Neuroimage. 2012; 60:623–632. [PubMed: 22233733]

Shehzad Z, Kelly AM, Reiss PT, Gee DG, Gotimer K, Uddin LQ, Lee SH, Margulies DS, Roy AK, Biswal BB, Petkova E, Castellanos FX, Milham MP. The resting brain: unconstrained yet reliable. Cereb Cortex. 2009; 19:2209–2229. [PubMed: 19221144]

Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull. 1979; 86:420–428. [PubMed: 18839484]

Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TE, Johansen-Berg H, Bannister PR, De Luca M, Drobnjak I, Flitney DE, Niazy RK, Saunders J, Vickers J, Zhang Y, De Stefano N, Brady JM, Matthews PM. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage. 2004; 23 (Suppl 1):S208–S219. [PubMed: 15501092]

Sørensen T. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. Biol Skr. 1948; 5:1–34.

Talairach, J.; Tournoux, P. Co-Planar Stereotaxic Atlas of the Human Brain. Thieme Medical Publishers, Inc; New York: 1988.

Thomason ME, Dennis EL, Joshi AA, Joshi SH, Dinov ID, Chang C, Henry ML, Johnson RF, Thompson PM, Toga AW, Glover GH, Van Horn JD, Gotlib IH. Resting-state fMRI can reliably map neural networks in children. Neuroimage. 2011; 55:165–175. [PubMed: 21134471]

Van Dijk KR, Hedden T, Venkataraman A, Evans KC, Lazar SW, Buckner RL. Intrinsic functional connectivity as a tool for human connectomics: theory, properties, and optimization. J Neurophysiol. 2010; 103:297–321. [PubMed: 19889849]

Van Dijk KR, Sabuncu MR, Buckner RL. The influence of head motion on intrinsic functional connectivity MRI. Neuroimage. 2012; 59:431–438. [PubMed: 21810475]

Weissenbacher A, Kasess C, Gerstl F, Lanzenberger R, Moser E, Windischberger C. Correlations and anticorrelations in resting-state functional connectivity MRI: a quantitative comparison of preprocessing strategies. Neuroimage. 2009; 47:1408–1416. [PubMed: 19442749]

Woolrich MW, Jbabdi S, Patenaude B, Chappell M, Makni S, Behrens T, Beckmann C, Jenkinson M, Smith SM. Bayesian analysis of neuroimaging data in FSL. Neuroimage. 2009; 45:S173–S186. [PubMed: 19059349]

Yan C, Liu D, He Y, Zou Q, Zhu C, Zuo X, Long X, Zang Y. Spontaneous brain activity in the default mode network is sensitive to different resting-state conditions with limited cognitive load. PLoS One. 2009; 4:e5743. [PubMed: 19492040]

Zhang Y, Brady M, Smith S. Segmentation of brain MR images through a hidden Markov random field model and the expectation–maximization algorithm. IEEE Trans Med Imaging. 2001; 20:45–57. [PubMed: 11293691]

Zuo XN, Xu T, Jiang L, Yang Z, Cao XY, He Y, Zang YF, Castellanos FX, Milham MP. Toward reliable characterization of functional homogeneity in the human brain: preprocessing, scan duration, imaging resolution and computational space. Neuroimage. 2013; 65:374–386. [PubMed: 23085497]

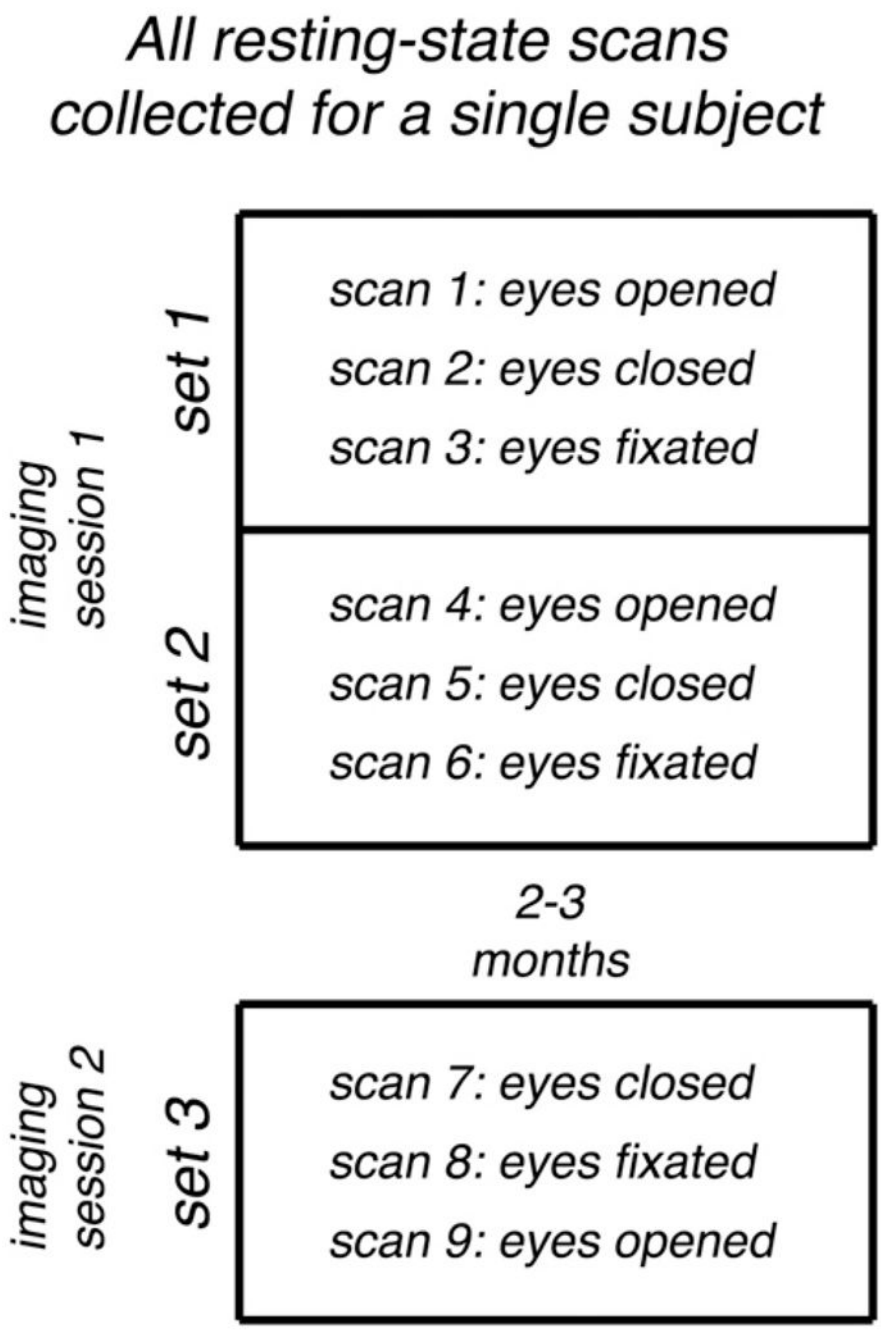

Fig. 1.

The collection of nine 10-minute resting-state fMRI scans with three different resting conditions, the order of which was counterbalanced across both subjects and sessions, for a single subject.

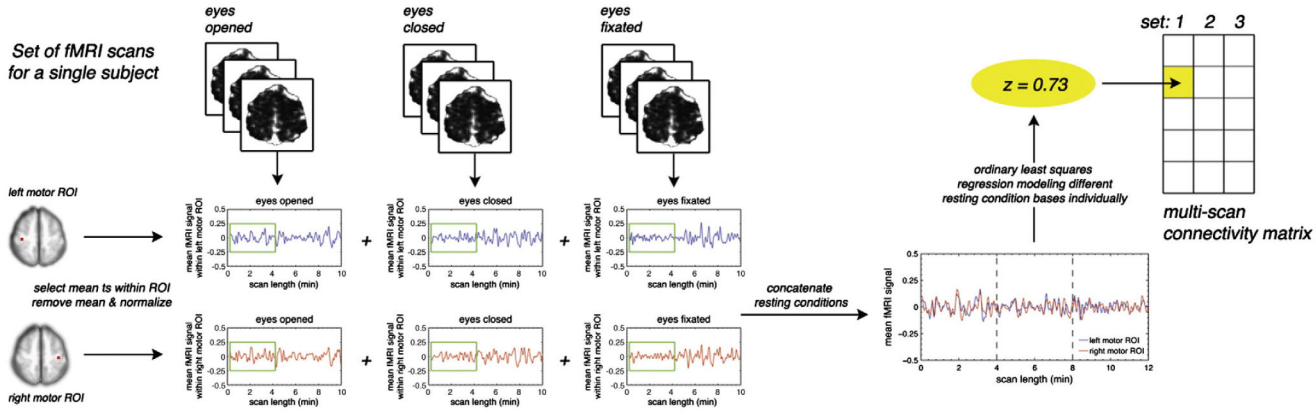

Fig. 2.

A schematic of the creation of scan lengths from the three resting conditions. Specifically, an example is given for the computation of functional connectivity between the left and right motor cortices for a scan length of 12 min. This process would be repeated for each of the different scan lengths.

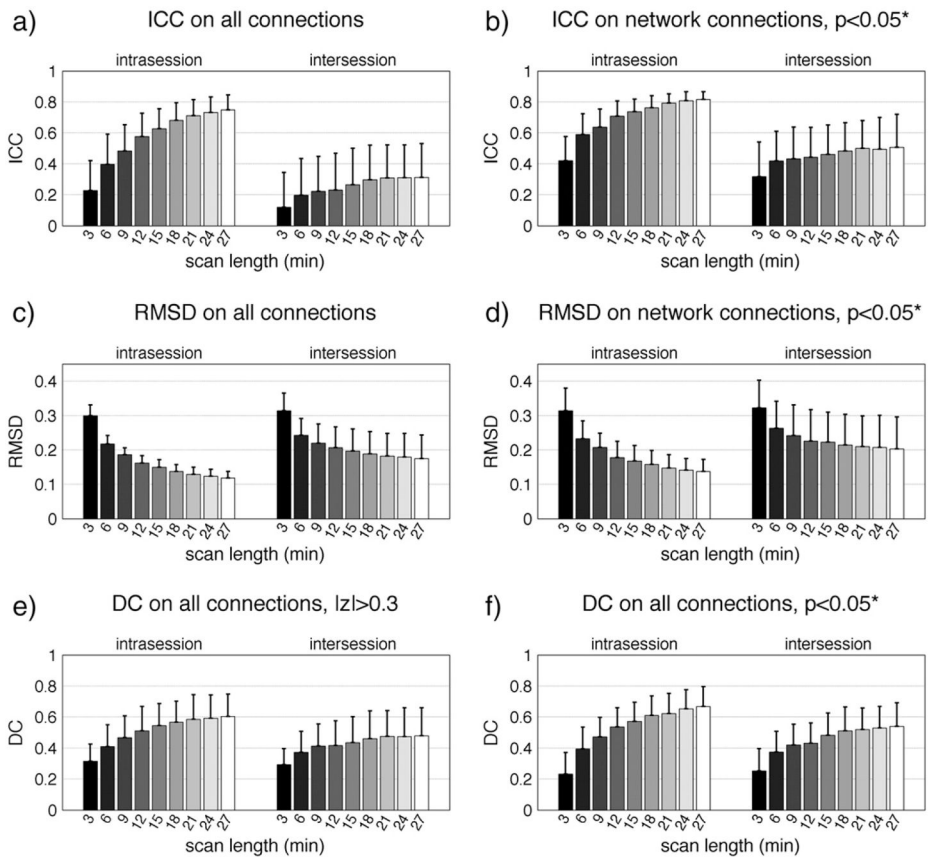

Fig. 3.

The intrasession and intersession intraclass correlation coefficients (ICCs), root-mean-square deviations (RMSDs), and Dice coefficients (DCs) are shown for the nine scan lengths. Bars represent means across specified group of connections (ICC) or subjects (RMSD and DC), and error bars are the standard deviations. The * denotes a Bonferroni-corrected p-value.

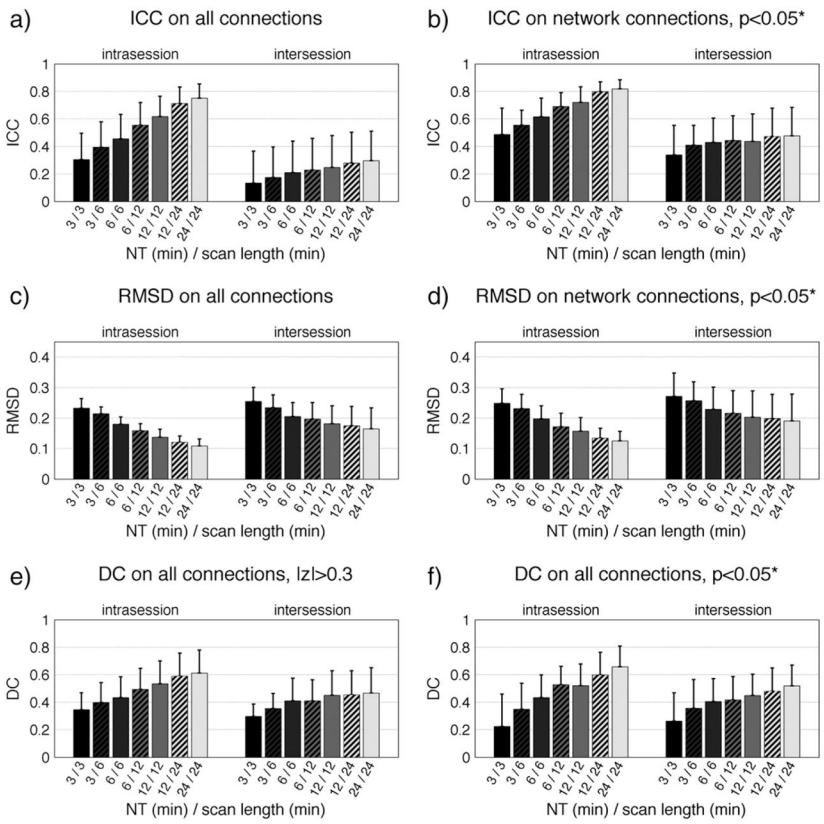

Fig. 4.

The intrasession and intersession ICCs, RMSDs, and DCs are shown for the scan lengths of 3, 6, 12, and 24 min as well as scan lengths of 6, 12, and 24 min with half the number of time points obtained by sampling every second time point. The numerator indicates the amount of time (in minutes) necessary to acquire the number of time points (NT) that were taken from the total imaging duration (in minutes) given by the denominator. Bars represent means across the specified group of connections (ICC) or subjects (RMSD and DC), and error bars are the standard deviations. The * denotes a Bonferroni-corrected p-value. Note that the time series used in the computation of FC for this comparison were not temporally filtered.

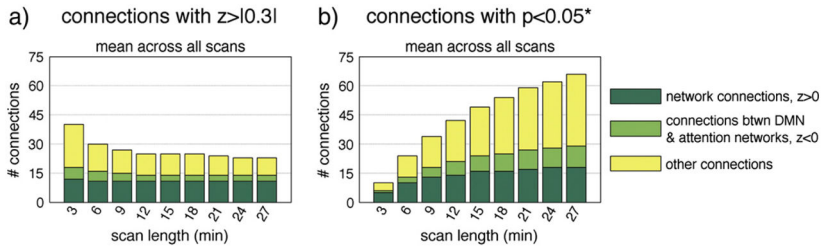

Fig. 5.

Mean number of connections with $|z| > 0.3$ and $p < 0.05$* over all scans are shown for the nine scan lengths. The * denotes a Bonferroni-corrected p-value.

Table 1

p-Values from one-way ANOVAs on scan length (categories: 3, 6, 9, 12, 15, 18, 21, 24, or 27 min). Repeated measures ANOVAs were used on RMSD and DC measures.

Effect of scan length	Intrasession	Intersession
All 153 connections		
ICC	7.38E-246	5.42E-18
RMSD	1.53E-114	1.22E-72
DC, $\|z\| > 0.3$	1.05E-42	3.72E-22
DC, $p < 0.05$*	1.05E-59	5.33E-37
*Network connections, $p < 0.05$**		
ICC	3.06E-30	0.108
RMSD	1.37E-61	1.77E-38

* denotes a Bonferroni corrected p-value.

Table 2

p-Values from two-way ANOVAs on the number of time points (NT) and scan length (categories: 3, 6, 12, or 24 min). Repeated measures ANOVAs were used on RMSD and DC measures. Note that the time series used in the computation of FC for this comparison were not temporally filtered.

Effect of length & NT	Intrasession		Intersession	
	Length	NT	Length	NT
All 153 connections				
ICC	4.20E-17	4.91E-05	0.212	0.426
RMSD	3.76E-10	9.39E-16	3.53E-04	2.13E-08
DC, $\|z\| > 0.3$	2.16E-04	0.051	0.135	0.036
DC, $p < 0.05$*	7.18E-05	0.051	0.114	0.264
*Network connections, $p < 0.05$**				
ICC	0.015	0.339	0.645	0.989
RMSD	1.78E-05	2.73E-05	0.111	1.33E-03

*denotes a Bonferroni corrected p-value.