



Published in final edited form as:

Neuroimage. 2021 October 01; 239: 118284. doi:10.1016/j.neuroimage.2021.118284.

Intra-session test-retest reliability of functional connectivity in infants

Yun Wang^a, Walter Hinds^a, Cristiane S Duarte^a, Seonjoo Lee^{a,b}, Catherine Monk^c, Melanie Wall^{a,b}, Glorisa Canino^d, Ana Carolina C. Milani^e, Andrea Jackowski^f, Marina Griorgi Mamin^g, Bernd U. Foerster^e, Jay Gingrich^a, Myrna M Weissman^{a,c}, Bradley S. Peterson^h, David Semanek^a, Edna Acosta Perez^{d,k}, Eduardo Labat^d, Ioannisely Berrios Torres^l, Ivaldo Da Silvaⁱ, Camila Parenteⁱ, Nitamar Abdala^j, Jonathan Posner^{a,*}

^aDepartment of Psychiatry, Columbia University Medical Center, New York, NY, USA

^bMental Health Data Science, New York State Psychiatric Institute, New York, NY, USA

^cDepartment of Obstetrics and Gynecology, New York State Psychiatric Institute, New York, NY, USA

^dSchool of Medicine, Medical Science Campus, University of Puerto Rico, San Juan, PR, USA

^eDepartment of Psychiatry, Federal University of Sao Paulo, Sao Paulo, Brazil

^fInterdisciplinary Lab for Clinical Neurosciences, Federal University of Sao Paulo, Sao Paulo, Brazil

^gDepartment of Pediatric, Federal University of Sao Paulo, Sao Paulo, Brazil

^hInstitute for the Developing Mind, The Saban Research Institute, Children's Hospital Los Angeles, CA, USA

ⁱDepartment of Gynecology, Federal University of Sao Paulo, Sao Paulo, Brazil

^jDepartment of Diagnostic Radiology, Federal University of Sao Paulo, Sao Paulo, Brazil

^kGraduate School of Public Health, Medical Science Campus, University of Puerto Rico, San Juan, PR, USA

^lBehavioral Science Research Institute, Academic Deanship, Medical Science Campus, University of Puerto Rico, San Juan, PR, USA

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

*Corresponding author. Jonathan.Posner@nyspi.columbia.edu (J. Posner).

Credit author statement

YW and JP conceptualized the study. CD, CM, JP, GC, AM, MM, BF, JG, MW, BP, EP, EL, IT, IS, CP, and NA collected the data. DS and WH organized, censored and preprocessed the data. YW and WH analyzed the data. SL and MW guided the statistical analysis. YW, WH, and JP wrote the manuscript. JP and YW reviewed and edited the manuscript.

Data and Code Availability Statements

The data that support the findings of this study are not publicly available due to the distinct regulations of IRBs for four studies from different countries (Brazil and US), but are available from the corresponding author on reasonable request.

The code used in this paper is publicly available at GitHub: https://github.com/wangyuncolumbia/Infant_finer_segmentation.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2021.118284.

Abstract

Resting functional MRI studies of the infant brain are increasingly becoming an important tool in developmental neuroscience. Whereas the test-retest reliability of functional connectivity (FC) measures derived from resting fMRI data have been characterized in the adult and child brain, similar assessments have not been conducted in infants. In this study, we examined the intra-session test-retest reliability of FC measures from 119 infant brain MRI scans from four neurodevelopmental studies. We investigated edge-level and subject-level reliability within one MRI session (between and within runs) measured by the Intraclass correlation coefficient (ICC). First, using an atlas-based approach, we examined whole-brain connectivity as well as connectivity within two common resting fMRI networks - the default mode network (DMN) and the sensorimotor network (SMN). Second, we examined the influence of run duration, study site, and scanning manufacturer (e.g., Philips and General Electric) on ICCs. Lastly, we tested spatial similarity using the Jaccard Index from networks derived from independent component analysis (ICA). Consistent with resting fMRI studies from adults, our findings indicated poor edge-level reliability (ICC = 0.14–0.18), but moderate-to-good subject-level intra-session reliability for whole-brain, DMN, and SMN connectivity (ICC = 0.40–0.78). We also found significant effects of run duration, site, and scanning manufacturer on reliability estimates. Some ICA-derived networks showed strong spatial reproducibility (e.g., DMN, SMN, and Visual Network), and were labelled based on their spatial similarity to analogous networks measured in adults. These networks were reproducibly found across different study sites. However, other ICA-networks (e.g. Executive Control Network) did not show strong spatial reproducibility, suggesting that the reliability and/or maturational course of functional connectivity may vary by network. In sum, our findings suggest that developmental scientists may be on safe ground examining the functional organization of some major neural networks (e.g. DMN and SMN), but judicious interpretation of functional connectivity is essential to its ongoing success.

Keywords

Infant; fMRI; Test-retest reliability; ICC; Subject-level; Edge-level; Network; Jaccard Index; ICA

1. Introduction

Providing a window into early brain development with limited influence of the post-natal environment, resting functional MRI (fMRI) studies of the infant brain have become an important tool for advancing developmental neuroscience (Duarte et al., 2020; Posner et al., 2016). Subtle changes in brain function early in development may, for example, confer lifelong implications on cognition, behavior, and psychopathology. Resting fMRI studies detail the organization of neural circuits based on functional connectivity, or the coherence of neural activity across disparate brain regions. In infant research, this approach may help identify early brain-based biomarkers, providing, for example, indicators of aberrant neurodevelopment, prenatal exposure effects, or neurobiological correlates of early cognitive functions, and may thereby facilitate early identification of at-risk individuals (Lugo-Candelas et al., 2018; Posner et al., 2016; Rudolph et al., 2018). Despite this promise and the growing use of resting fMRI studies in infant research, the reliability of functional

connectivity measures in the infant brain remains largely unexamined. Establishing test-retest reliability is critical to solidifying the methodological rigor and basis for including resting fMRI techniques within the armamentarium of neurodevelopmental researchers.

Prior studies have examined the intra-session (so-called short-term) and between-session test-retest reliability of functional connectivity measures in the adult and child brain (Marchitelli et al., 2016; Noble et al., 2019; Somandepalli et al., 2015). Most studies have used the Intraclass correlation coefficient (ICC) as the primary measure of test-retest reliability for functional connections due to the flexibility of the ICC to explicitly model known sources of measurement variance (Shrout and Fleiss, 1979; Somandepalli et al., 2015). Across a wide range of approaches, these studies have typically found poor reliability for edge-level ICCs (Noble et al., 2019), yet moderate-to-good reliability for subject-level ICCs (Mejia et al., 2018; Mejia et al., 2015; Somandepalli et al., 2015). In the context of functional connectivity, an “edge-level” ICC refers to the test-retest reliability of a single functional connection (an edge). For every pairwise set of edges between regions-of-interest (ROIs) across the brain, an ICC can be calculated resulting in thousands of ICCs per brain scan. In contrast, a “subject-level” ICC is calculated across a connectivity matrix - that is, multiple connections within one brain scan, or within one brain network, are considered collectively. The test-retest reliability of both edge- and subject-level ICCs merit consideration because: 1) edge-level connectivity is the most basic element in functional connectivity analysis, and not surprisingly, is often used in fMRI research; and 2) subject-level connectivity is increasingly used to index patterns of brain connectivity, for instance in graph theory and fingerprinting (Finn et al., 2015; Gordon et al., 2017).

Because of the temporally correlated nature of time series fMRI data, the time interval between scanning sessions (i.e., short-term versus long-term) is an important consideration when assessing reliability. Several studies, for example, have found repeated measures taken over shorter intervals (e.g. intra-session or within days) are more reliable than those taken over longer intervals (e.g. weeks to years). Herein, we examine intra-session test-retest reliability for two primary reasons. First, intra-session reliability has not yet been established in infants. Second, long-term reliability in infants may be confounded by neurodevelopmental changes that occur rapidly in infancy.

Resting fMRI reliability estimates may be influenced by MRI acquisition and processing procedures. For example, studies have reported differences in connectivity depending on the MRI scanning site (Biswal et al., 2010; Jovicich et al., 2016). These effects are detected even after harmonization techniques are conducted (Noble et al., 2017). Also, reliability improves with the acquisition of more frames of fMRI data per subject either by increasing the duration of the scanning sessions or by increasing the number of runs (Birn et al., 2013). Lastly, several studies have found that pre-processing procedures designed to reduce the contribution of noise, such as motion, global signal, white matter, and cerebrospinal fluid, to fMRI signal can have unexpected effects on reliability estimation (Guo et al., 2012; Parkes et al., 2018; Shirer et al., 2015). For example, motion correction, a technique used to reduce the influence of head motion on fMRI signal, can, somewhat counterintuitively, *reduce* test-retest reliability when the motion itself is consistent over time (Parkes et al., 2018). Of note,

ROIs from different atlases can vary substantially (Salehi et al., 2020), resulting in different ICCs.

Despite the numerous test-retest reliability studies that have been conducted in adults and children, prior studies have not reported reliability estimates in infants, nor have they examined test-retest reliability across infant studies conducted at different scanning sites. The moderate-to-good reliability for subject-level ICCs reported in studies of children and adults (Marchitelli et al., 2016; Noble et al., 2017; Somandepalli et al., 2015) is not sufficient to assume similar reliability will also be present in infants. First, significant neurovascular changes occur throughout infancy, even over days, and these changes may influence neurovascular coupling, introducing noise, and potentially diminishing the reliability of functional connectivity estimates (Kozberg and Hillman, 2016). Second, resting fMRI studies of infants are conducted during sleep (without sedation), whereas studies in adults and children are more commonly done while participants are awake. Different sleep states may be associated with distinct connectivity patterns, yet sleep state is seldom assessed or controlled for in infant fMRI research (Lee et al., 2020). Third, the increased water content of the infant brain, relative to a more mature brain, alters tissue contrast (e.g., T2-weighted images in infants provide clearer grey-white matter demarcation) and may make image processing more challenging, eroding measurement reliability. Fourth, functional networks in infants may be in a premature state, and potentially less stable, rendering their measurement less reliable.

Our primary objective in this paper is to investigate the short-term (within one MRI scanning session) test-retest reliability of functional connectivity measures in the newborn infant brain across four different studies, each conducted at a different site (Table 1). First, we investigate edge-level and subject-level test-retest reliability measured by ICCs from ROI-based whole-brain connectivity, as well as from two resting fMRI networks commonly investigated by ROI analysis - the default mode network (DMN) and the sensorimotor network (SMN). Second, we examine the influence on ICCs of run duration, study site, and scanner manufacturer (e.g., Philips and General Electric). Lastly, we test the spatial reproducibility from networks derived from independent component analysis (ICA), a data-driven method to assess functional connectivity that is distinct from ROI-based approaches. For these ICA-based analyses, we compare the ICA-derived network across runs (and term this, *spatial reproducibility*), as well as with canonical ICA-derived networks from adults (and termed this, *spatial similarity*).

2. Method

2.1. Datasets and infant preparation protocol

We conducted this secondary data analysis from a sample of 239 infants acquired as part of four neurodevelopmental studies. All study procedures were approved by their respective Institutional Review Boards. In each of these studies, infants were scanned during natural sleep, without sedation using an infant preparation protocol developed at the New York State Psychiatric Institute (NYSPI) over a decade ago for unsedated infant research MRI scans. With this "natural rhythm" approach, we prepare the infant for the MRI scan using the infant's normal pre-nap routine (feeding, changing, rocking, swaddle), and the family is

encouraged to align the infant's schedule to the scan time in the days before the scan. Infants are also attenuated to the MRI scanner noises either at home, or during the preparation process on the scan day, or both. The MRI pulse/scanning sequences and inclusion/exclusion criteria for each study are described in the Supplementary Materials 1.1 and 1.2; respectively.

After quality checking (see 2.2 for details), we included 119 infants in our analyses. Table 1 summarizes the demographics for these 119 infants, including sex and postmenstrual age (PMA) at scan, as well as the scanning parameters used for each study. It is worth noting that the infant age range/variability within each study was limited, yet statistically significant differences in age across studies were detected. Therefore, we adjusted for age in the linear regression analysis (see Section 2.4.2).

2.2. Data analytic plan

As outlined in Fig. 1, our data analytic plan consisted of four components: data curation, data processing, connectivity estimates, and test-retest reliability.

2.2.1. Data curation—We aggregated infant MRI datasets across the four neurodevelopmental studies yielding 239 infant scans. We conducted an initial quality assessment of the T2-weighted and resting fMRI scans by trained technicians who visually inspected the scans and removed those with incomplete data (e.g., no T2w scan and/or severe imaging artifacts, such as ghosting). A total of 90 out of 239 scans (~37.6%) were excluded on this basis. Supplementary Table 1 indicates the number of scans excluded for each study.

2.2.2. Data processing

2.2.2.1. Scrubbing. For the resting fMRI scans, frames (or volumes) corrupted by head motion were marked as outliers and removed - a process called “scrubbing” (Mongerson et al., 2017). We adopted a common criterion for identifying outlier frames - a combination of frame-wise displacement (FD) and blood-oxygen level dependent (BOLD) data variance (DVARS). DVARS is an estimate of framewise changes in BOLD signal intensity and is calculated as the root-mean-squared variance of the temporal derivative of all brain voxel time series (Power et al., 2012). As suggested elsewhere (Power et al., 2013), frames were marked as outliers and removed if either of the following were true: $FD > 0.2$ mm or $DVARS > 3\%$.

We excluded entire functional runs if 1) there was only one run per subject, and the total scanning length after scrubbing was less than 5 min; OR 2) there was more than one run per subject, but the run length after scrubbing was less than 2.5 min. These criteria translated approximately to a combination of 2.5 min minimum per run after scrubbing and no more than 30–40% of scrubbed frames per run. Thirty scans were removed based on these criteria, leaving 119 usable scans for our analyses (see Supplementary Table 1). The percentage of outlier frames within these 119 usable scans is summarized in Supplementary Figure 1. We also compared head motion after scrubbing, based on frame-wise displacement (FD), across

our four study sites and did not detect significant site differences (see Supplementary Figure 2).

2.2.2.2. Pre-processing functional data.: After visual inspection and scrubbing, the functional scans were motion-corrected using MCFLIRT (Jenkinson et al., 2002), which transforms functional frames via rigid alignment and outputs an estimate of the six rigid motion parameters. We then implemented brain extractions (Smith, 2002), used a high-pass filter with a 100-second period cutoff value, and a low-pass spatial filter using FSL's SUSAN, and a kernel of 4.5 millimeters for spatial smoothing (Smith and Brady, 1997). Finally, we scaled the mean signal intensity to 10,000.

2.2.2.3. Nuisance regression.: To reduce physiological noise due, for example, to cardiac and respiratory rhythms, and to minimize the influence of head motion, we regressed the whole-brain time course against the mean global signal, signal extracted from white matter (WM) and ventricular masks, as well as motion parameters (see Sect 2.2.2.b) and their derivatives. This approach to signal correction has been shown to provide a good trade-off between maintaining degrees of freedom and removing noise contamination (Ciric et al., 2017).

2.2.2.4. T2-weighted segmentation and normalization.: First, we used the Developmental Human Connectome Project (dHCP) automated pipeline for segmentation of T2w scans (Makropoulos et al., 2018). Second, for finer anatomical segmentation by lobe, we then further subdivided grey matter from the frontal, parietal, and occipital lobes with a custom algorithm that utilizes a warped segmentation from the individual subject to the University of North Carolina (UNC) AAL neonatal atlas (Shi et al., 2011). The estimated warp, or "normalization," was carried out with Advanced Normalization Tools (ANTs). All normalization results were visually inspected to confirm accuracy.

Our hybrid structural atlas combined dHCP segmentations (87 regions) with the UNC AAL neonatal atlas (90 regions). Whereas the dHCP pipeline generates individualized segmentations, it does not define subregions of the frontal, parietal, or occipital lobes. Conversely, the UNC AAL neonatal atlas provides finer parcellation of frontal, parietal and occipital lobes, but defines regions at the group level which may obscure individual anatomical characteristics. Taken together, we created a hybrid atlas using the dHCP segmentations and then further subdivided the frontal, parietal, or occipital lobes based on the AAL neonatal atlas. This hybrid atlas generated 101 ROIs with whole-brain coverage, as listed in Supplementary Table 2. Our hybrid atlas and the related code are publicly available at GitHub: https://github.com/wangyuncolumbia/Infant_finer_segmentation.

2.2.2.5. Co-registration and normalization.: We calculated a mean functional volume and used this as the moving volume in an affine registration with the segmented structural scan, based on a normalized mutual-information (NMI) cost function in FLIRT (Smith et al., 2004). We then used this initial registration to guide a subsequent, and more accurate, boundary-based registration (BBR), also implemented with FLIRT, which utilizes the grey-white boundary to calculate the cost function (Greve and Fischl, 2009). Because averaging functional data can blur the grey-white boundary, we used an exemplar volume from the

middle of the functional run for this BBR step. Lastly, we normalized the functional scans by applying the individual's structural-to-template warp. To reduce the number of data interpolations to a single instance, we applied all linear transformations and the non-linear warp simultaneously through `antsApplyTransforms` (ANTs v2.1.0) using Lanczos interpolation.

2.3. Connectivity estimates

2.3.1. Connectivity matrix based on a structural atlas—We derived a whole-brain ROI-to-ROI connectivity matrix based on our individualized segmentation of 101 ROIs - that is, we calculated a 101×101 connectivity matrix based on the Pearson correlation of the averaged time series fMRI data from each ROIs (Supplementary Table 2). Pairwise connections were then normalized by Fisher's *r*-to-*z* transformation. We also generated ROI-to-ROI connectivity matrices for two resting fMRI networks - the DMN, the SMN, each with 10 ROIs (see Supplementary Table 2 *colored labels*).

2.3.2. Connectivity matrix based on a functional atlas—To explore whether the reliability of connectivity estimates was affected by the approach used to parcellate the brain, we also derived a whole-brain ROI-to-ROI connectivity matrix based on a predefined functional atlas, a.k.a. the Power atlas (Power et al., 2011), in parallel with our approach using a structural atlas (see 2.3.1). We excluded from the Power atlas nodes from domains not relevant to our reliability analyses, including nodes in the “Uncertain” or “Cerebral White Matter” categories (Power et al., 2011). In total, 216 ROIs were included. We then warped and co-registered these 216 spherical ROIs (radius=3mm) to our infant structural and functional MRI scans. Registration from the adult template to the infant space were visually inspected. If these were of poor quality, we used the following procedures: first, we used ITK-SNAP (Yushkevich et al., 2006) to visually guide manual alignment and optimize the initial transformation matrix; then we ran `antsRegistration` and `antsApplyTransforms` to register the images again with this improved initial transformation matrix. The final registration was again visually inspected. Finally, we derived a whole-brain 216×216 connectivity matrix, and normalized pairwise connections using Fisher's *r*-to-*z* transformation to allow for comparisons.

2.3.2.a. Hierarchical cluster analysis. Whereas the Power Atlas denotes several functional networks, prior infant research suggests that only a few functional networks (e.g. the sensorimotor and auditory networks) are formed by infancy (Gao et al., 2015). To test this in our data, we conducted a hierarchical cluster analysis from our group-average functional atlas-based connectivity map. The resultant cluster map suggested that the default mode network (DMN), sensorimotor, and auditory networks grouped together (see Supplementary Figure 3). Other networks, e.g., salience, and dorsal attention network, failed to form continuous clusters. Therefore, in our functional atlas-based analysis, we limited our analyses to the DMN, sensorimotor, and auditory networks.

2.3.3. Independent component analysis (ICA)-based connectivity—We used MELODIC to perform group-level independent component analysis (ICA) separately for each run and for each study (Smith et al., 2004). The number of components was set at 20

for our main analysis but varying the number of components from 15~75 yielded similar results (see Supplementary Figure 5). To accurately identify and label resting fMRI networks extracted from the group ICA, we compared the components to well-defined resting fMRI networks obtained using similar ICA procedures in adults (Smith et al., 2009). Using ITK-SNAP (Yushkevich et al., 2006) and Advanced Normalization Tools (ANTs) (Avants et al., 2009), we registered and warped the 20 components derived from our infant group-ICA to a template infant atlas (Shi et al., 2011), and then similarly registered 20 components from an adult group-ICA (Smith et al., 2009) to the same template space. We then calculated the Jaccard index (see 2.4 for explanation of Jaccard index) between the infant and adult components and labeled the infant components based on their highest Jaccard index with a corresponding adult component (e.g., the infant ICA component showing the highest Jaccard index with the adult DMN was labeled the infant DMN component). In addition, to examine spatial reproducibility across runs, we performed dual regression (Nickerson et al., 2017) to obtain subject-specific sets of spatial maps for each run and each subject. To binarize the data, we thresholded the spatial maps at $|z| > 2.3$. Lastly, to compare the spatial similarity of the networks identified across our study sites, we aggregated the functional data across runs from each study site and conducted a separate group ICA for each of the four study sites, then compared them using the Jaccard index.

2.4. Test-retest reliability

2.4.1. Test-retest reliability from ROI-based analysis—We calculated the edge level and network level Intraclass correlation coefficient (ICC [2,1]) (Chen et al., 2018; Noble et al., 2019) using the following formula:

$$ICC = \frac{MS_R - MS_E}{MS_R + (k - 1)MS_E}$$

MS_R = mean square for rows; MS_E = mean square for error; k = number of raters/measurements. Rows for ICC calculation represent the number of subjects or observations.

For Studies 1 and 2, which both included two or more runs, we compared the functional connectivity estimates between the two runs (Study 1) and between earlier runs to sequentially subsequent runs (Study 2), henceforth, referred to as “*between-run reliability*” (Fig. 2A). Because Studies 3 and 4 included only one run, we could not calculate between-run reliability. Instead we compared functional connectivity estimates from the first half of the run with those from the second half - henceforth, referred to as “*within-run reliability*” (Fig. 2B).

For between-run and within-run reliability, we examined both edge-level and subject-level ICCs. To allow for site comparisons, we also calculated within-run reliability for Study 1. (We did not do this for Study 2 because the runs were only 75 frames, too short to estimate within-run reliability). The interpretation of the ICC is commonly categorized as follows: poor <0.4, fair 0.4–0.59, good 0.6–0.74, excellent 0.75 (Noble et al., 2019).

2.4.2. Effects of run duration, site, and MRI manufacturer—We next examined the influence of run duration, study site, and MRI manufacturer on test-retest reliability. Because of collinearity between study site and MRI manufacturer, we examined the effect of study site and MRI manufacturer on “within-run reliability” in separate linear regression models, adjusting for age, sex, and head motion (as index by FD, see Sect 2.2). Furthermore, we conducted multiple linear regression to test for effects of run duration and study site (Studies 1 and 2 used the same type of MRI scanner) on “between-run reliability,” adjusting for age, sex, and head motion.

To construct different lengths of run duration, we used Study 1 because this study provided scans with two resting fMRI runs, each over 7 minutes in duration, which of our four studies provided the longest run duration. After scrubbing, we binned each scanning run into batches of frames, ranging from 75 frames to 150 frames in increments of 15 (Fig. 2C). For example, for a scan with 150 frames or more in each of two runs, the data were binned into two sets of 75 frames, two sets of 90 frames, etc., up to two sets of 150 frames. The frames used were the initial frames for each run - for instance, in the previous example, the two sets of 75 frames were comprised of the first 75 frames from the first run, and the first 75 frames from the second run.

2.4.3. Spatial reproducibility of ICA-derived networks—For our ICA-based analysis, we used the Jaccard index to calculate the spatial reproducibility of the derived networks across runs. The Jaccard index measures the similarity between two spatial network maps by computing the normalized amount of overlap; that is, the ratio of intersection relative to union of two spatial network maps. Jaccard Index ranges from 0 to 1; a high Jaccard index denotes high similarity of two spatial maps.

We limited our ICA-based analysis of spatial reproducibility to Study 1 because this study contained two runs allowing for comparison of the ICA-derived networks from the first run with those from the second. ICA estimates are not recommended for shorter runs (Glasser et al., 2018) and thus we excluded Study 2. Studies 3 and 4 were excluded because they contained only one run per infant.

2.4.4. Spatial similarity of ICA-based networks with adult networks—We used the Jaccard index to evaluate the spatial similarity of ICA-derived networks from our infant studies with well-defined resting fMRI networks obtained using similar ICA procedures in adults (Smith et al., 2009). We conducted these comparisons for five resting fMRI networks: the default mode network (DMN), the sensorimotor network (SMN), the visual network (VN), and the left and right executive control network (ECN) and conducted these comparisons for each of our four study sites separately. Comparisons with the additional components reported by Smith et al. (Smith et al., 2009) are provided in Supplementary Figure 7.

3. Results

3.1. Test-retest reliability from ROI-based analysis

Below we present the reliability results obtained from the structural atlas. The results from the functional atlas (the Power atlas) were similar and are presented in Fig. 3 with the related statistics provided in the Supplemental Material 1.3.

3.1.1. Edge-level ICCs—We analyzed the reliability of all possible pairwise combinations of 101 ROIs across the whole brain, yielding 5,050 ICCs per scan. Between-run reliability was poor for runs of 75 frames (mean ICC = .146, 95% CI [.142, .151]) and poor for runs of 150 frames (mean ICC = .179, 95% CI [.175, .184]; see Fig. 3). Within-run reliability was fair (mean ICCs = .41, 95% CI [.403, .410]; see Supplementary Figure 4).

Regression analysis of between-run reliability showed a significant effect ($F=81.95$, $p<10^{-10}$) of run duration ($t=12.64$, $p<10^{-10}$) and study site ($t=-3.27$, $p=10^{-3}$). Linear regression on within-run reliability also confirmed a significant effect of study site ($F=137.4$, $p<10^{-10}$) and MRI scanner ($F=7.79$, $p<.005$).

3.1.2. Subject-level ICCs

3.1.2.1. Subject-level ICCs - Whole-brain analysis.: Across the whole brain, between-run reliability was fair for runs of 75 frames (mean ICC = .43, 95% CI [.40, .47]) and fair-to-good for runs of 150 frames (mean ICC = .51, 95% CI [.46, .55]). Within-run reliability was good (mean ICCs = .63, 95% CI [.61, .65]). Regression analysis of between-run reliability showed a significant effect ($F=10.32$, $p<10^{-5}$) of run duration ($t=3.354$, $p=10^{-3}$) and study site ($t=-3.17$, $p=.002$). Neither significant effect of study site ($F=0.25$, $p=0.62$) nor MRI scanner ($F=1.78$, $p=0.19$) was found for within-run reliability.

3.1.2.2. Subject-level ICCs - Network-based analysis.: For the DMN, between-run reliability was good for runs of 75 frames (mean ICC = .63, 95% CI [.59, .68]) and good-to-excellent for runs of 150 frames (mean ICC = .73, 95% CI [.69, .78]). Within-run reliability was also excellent (mean ICCs = .76, 95% CI [.73, .78]). Regression analysis of between-run reliability showed a significant effect ($F=19.52$, $p<10^{-5}$) of run duration ($t=4.24$, $p<10^{-3}$) and study site ($t=-4.72$, $p<10^{-3}$). Regression analysis of within-run reliability showed a significant effect of MRI scanner ($F=14.50$, $p<10^{-3}$) but not study site.

For the SMN, between-run reliability was fair-to-good for runs of 75 frames (mean ICC = .60, 95% CI [.55, .66]) and good for runs of 150 frames (mean ICCs ranges from .69, 95% CI [.63, .74]). Within-run reliability was good-to-excellent (mean ICCs = .72, 95% CI [.69, .74]). Regression analysis on between-run reliability showed a significant effect ($F=11.93$, $p<10^{-5}$) of run duration ($t=3.22$, $p<0.002$), and study site ($t=-3.77$, $p<10^{-3}$). Regression analysis of within-run reliability showed a significant effect of MRI scanner ($F=12.62$, $p<10^{-3}$) but not study site.

3.2. Spatial reproducibility of ICA-derived networks

Using dual regression, we obtained subject-specific spatial maps for each of the two runs in Study 1 and identified five networks for comparison: DMN, SMN, VN, and the left and right executive control network (ECN). Based on the Jaccard index between runs, we found high spatial reproducibility of the DMN, SMN, and VN, but worse reliability for the left and right ECN (Jaccard indices: DMN=0.54, SMN=0.53, VN=0.54, left ECN=0.25, right ECN=0.23; Fig. 4B presents the distribution of Jaccard indices stratified by network). We then compared the ICA-derived networks in infants with those from adults and found moderate spatial similarity between infants and adults for the DMN, SMN, and VN, but not for the left or right ECN (Jaccard indices ranged from 0.43 to 0.65 for DMN, SMN, and VN; but from 0.17 to 0.20 for left and right ECN; Fig. 4A). Moreover, the Jaccard indices between the infant versus adult networks varied across sites; for example, for the DMN, it ranged from 0.64 for Study 1 to 0.39 for Study 2. As seen in Fig. 4C, the precuneus and posterior cingulate - posterior hubs of the DMN - were consistently observed within the infant DMN across all four studies, however, the medial prefrontal cortex - an anterior hub of the DMN - was only encompassed within the DMN component in Studies 1 and 3. In Supplementary Figure 5, we present comparisons of the SMN, VN, and the left and right executive control network (ECN) across our four study sites. In Supplementary Figure 6, we present comparisons of the SMN, VN, and the left and right executive control network (ECN).

4. Discussion

To our knowledge, this is the first study to investigate the short-term (intra-session) test-retest reliability of functional connectivity in infants across multiple sites and run durations. Three observations were noted. First, whereas individual connections (i.e. edge-level) exhibited poor reliability, subject-level reliability was moderate to high. Second, reliability improved as the number of frames increased, but reliability varied by study site and scanning manufacturer. Third, ICA-derived networks in infants exhibited moderate spatial reproducibility within scanning sessions and were spatially similar to adult networks; however, this was not true of all networks. Differences in spatial reproducibility across networks may indicate that networks differ in their maturational course with some developing earlier (e.g. DMN, SMN, and VN), and others later (e.g., left and right ECN). This developmental hypothesis could be more rigorously tested in subsequent longitudinal research. Similarly, the organization of networks, such as the DMN, may increase over the first few years of life (Grayson and Fair, 2017). For example, adult studies consistently demonstrate midline connections within the DMN with connectivity between posterior and anterior hubs including the posterior cingulate and medial prefrontal cortices, respectively. This posterior-anterior axis of the DMN was not consistently detected in our studies (Fig. 4C) suggesting that this connection may emerge later in development.

Our reliability findings were largely in line with those reported in adult and child studies. For example, Pannunzi et al. examined the test-retest reliability of functional connectivity in 50 healthy adults and reported low reliability for single connections with an average edge-level ICC of 0.22 (Pannunzi et al., 2018). Even with long scan durations of 30 min (Birn et al., 2013), edge-level ICCs remained limited (~0.4). A recent meta-analysis of 25 resting

fMRI studies encompassing over 2,000 participants concluded that individual connections have an average ICC of 0.29 (Noble et al., 2019). Conversely, when functional connectivity is examined at the subject level, reliability generally improves (Mejia et al., 2018; Mejia et al., 2015), with longer scanning lengths improving these results even further (Mejia et al., 2018). Recent advances, termed precision resting fMRI, suggest that high sampling rates coupled with extended scanning session (e.g. over 60 min), test-retest reliability can reach ICCs in the excellent range (0.8–0.9) (Gordon et al., 2017; Gratton et al., 2019). Additionally, the spatial reproducibility of ICA-derived networks have been consistently reported to be high in adults (Braun et al., 2012; Marchitelli et al., 2016). In sum, our reliability findings from infant resting fMRI data contrast poor edge-level reliability with good subject-level reliability and parallel the existing literature from adults.

Our reliability findings are simultaneously both reassuring and concerning and suggest the following guidelines when using resting fMRI techniques in infant neuroimaging research. First, developmental neuroscientists should take some measure of comfort in knowing that subject-level measures from infant resting fMRI data can be estimated reliably with relatively short scanning durations (e.g. 5–10 min). Second, our findings suggest caution against reliance upon measures of individual functional connections. This is unfortunate because an individual functional connection is the most basic unit of analysis in resting fMRI research, and its interpretation compared to subject-level measures is more straightforward. Nonetheless, several subject-level measures are available and can be used to conduct statistical comparisons or to test brain-behavior correlations - for example, higher-order graph theory metrics (network density and clustering) (Braun et al., 2012; Cao et al., 2014; Termenon et al., 2016; Wang et al., 2019), and dynamic functional measures (Choe et al., 2017). Third, we found that reliability varied across resting fMRI networks derived from ICA, scanning sites and by scanning manufacturer. These last two points underscore the importance of data harmonization for multi-site studies, although site differences may persist nonetheless (Noble et al., 2017). While an ideal approach to examine site effects would have been a travelling human phantom (Huynh et al., 2019), the feasibility of this approach would be limited in a sample of newborns. Fourth, we did not find significant associations between reliability and infant age at scan, suggesting for example, that connectivity measures can be compared across infancy, albeit within the limited age range examined in our studies (postmenstrual age 39 – 59 weeks). Fifth, our data quality procedures resulted in nearly 50% of scans being excluded. Less rigorous data quality procedures may impact reliability, and potentially reduce measurement validity. This high level of data loss underscores the difficulty in obtaining high quality infant MRI data. Allowing ample time to prepare infants for the MRI scan (e.g. feeding and swaddling) and to fall asleep is imperative for successful data collection.

Limitations of our study are important to consider. First, none of the four studies in our analysis used multiband imaging, which can significantly increase sampling rates. Increasing the rate of data acquisition, multiband imaging might improve the reliability of infant functional connectivity measures, although some research suggests that reliability is more affected by scan duration than sampling rate (Birn et al., 2013; Demetriou et al., 2018). Second, we used the Developmental Human Connectome Project automated pipeline (Makropoulos et al., 2018) to segment the infant brains in our study, and added additional

parcellations to subdivide the frontal, parietal, and occipital lobes (see 2.2.2). We also used a functional atlas (Power et al., 2011) that yielded similar reliability estimates (see 2.3.2). However, other approaches to infant brain segmentation are available (e.g. the newly released Infant FreeSurfer pipeline) and merit investigation. The approach used for segmentation can influence the reliability of functional connectivity estimates - for example, adult resting MRI research indicates that more accurate segmentation improves functional connectivity measures (Salehi et al., 2020). To date, however, comparisons of tools for infant brain segmentation are scarce (Hashempour et al., 2019). Moreover, the infant brain is highly dynamic, increasing substantially in size and organization over development. As a result, for any given segmentation tool, its accuracy may exceed others at one time point in infancy, and yet become less accurate at subsequent time points. Brain segmentation with artificial intelligence (AI), by affording accuracy, speed, and flexibility, may offer a path forward, maintaining precision despite the dynamism of the infant brain. AI-based tools are still under development and will require validation (Wang, 2020). Third, we adopted a standard procedure to scrub motion outliers and further corrected motion with MCFLIRT in FSL. However, other types of motion correction techniques such as ICA-FIX or ICA-AROMA might change the reliability (Marchitelli et al., 2016; Parkes et al., 2018). Preprocessing procedures such as spatial smoothing could influence reliability, particularly given the small volume of the infant brain. Likewise, criteria for identifying head motion outliers, such as thresholds for DVARS and FD, need to be validated in infants.

Three additional methodological factors merit further consideration. First, while we detected site differences in our reliability estimates, there were demographic differences across our samples, including race, ethnicity, and other maternal characteristics, as well as differences in the MRI pulse sequences (see Supplementary Material 1.1). We therefore cannot determine the cause of these site differences in reliability estimates. At the same time, it is important to bear in mind that the diversity of our sample (e.g. four sites, two continents, two countries, and one commonwealth) should also be considered a strength - a more homogeneous sample would raise concerns of generalizability. For example, if our study included only one site with one relatively homogenous population, then it would remain unclear whether our findings generalize to other populations. Second, we used 8-channel head coils across the four study sites in this work. Although no study has directly compared FC reliability across head coils that differ in their number of channels (e.g., 8 vs. 32 channels), we speculate that reliability may increase with higher-quality fMRI data (Panman et al., 2019). Yet, data suggest that as the number of channels increases, inhomogeneous distribution of signal-to-noise ratios across the whole brain also increases with disparate effects on regional FC reliability estimates (Wiggins et al., 2006). Future studies are needed to further investigate this issue. Third, we did not assess infant sleep state during scanning which could impact functional networks and their reliability, especially for networks involving frontal areas, e.g., ECN and DMN. In adult studies, researchers have shown sleep state reduces the involvement of the frontal cortex in relation to the DMN (Horovitz et al., 2009); and recent infant work has shown that higher-level cognition and related brain networks are particularly affected by sleep state (Eyre et al., 2021; Linke et al., 2018; Raz and Saxe, 2020). Although we found inconsistent integration of frontal areas with the DMN across sites, our study cannot answer whether this variable integration is related to infant

sleep state or to incomplete maturation of the DMN. Future studies might elucidate this question by collecting infant fMRI data with simultaneous sleep state data (e.g., EEG). Similarly, it is possible that in addition to brain maturation, sleep state could help explain the lower reproducibility that we found for the ECN.

In sum, our study may offer useful parameters for the use of infant resting fMRI as a methodology for neurodevelopmental research. For example, whereas caution should be taken against interpreting behavioral associations, or the predictive validity, of specific functional connections, developmental scientist may be on firmer ground when associating the functional organization of networks (e.g. DMN and SMN) with neurodevelopmental outcomes. Infant resting fMRI research is rapidly advancing, and judicious interpretation of its measures is essential to its ongoing success.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

- Avants BB, Tustison N, Song G, 2009. Advanced normalization tools (ANTS). *Insight j* 2 (365), 1–35.
- Birn RM, Molloy EK, Patriat R, Parker T, Meier TB, Kirk GR, Nair VA, Meyerand ME, Prabhakaran V, 2013. The effect of scan length on the reliability of resting-state fMRI connectivity estimates. *Neuroimage* 83, 550–558. [PubMed: 23747458]
- Biswal BB, Mennes M, Zuo XN, Gohel S, Kelly C, Smith SM, Beckmann CF, Adelstein JS, Buckner RL, Colcombe S, Dogonowski AM, Ernst M, Fair D, Hampson M, Hoptman MJ, Hyde JS, Kiviniemi VJ, Kötter R, Li SJ, Lin CP, Lowe MJ, Mackay C, Madden DJ, Madsen KH, Margulies DS, Mayberg HS, McMahon K, Monk CS, Mostofsky SH, Nagel BJ, Pekar JJ, Peltier SJ, Petersen SE, Riedl V, Rombouts SARB, Rypma B, Schlaggar BL, Schmidt S, Seidler RD, Siegle GJ, Sorg C, Teng GJ, Veijola J, Villringer A, Walter M, Wang LH, Weng XC, Whitfield-Gabrieli S, Williamson P, Windischberger C, Zang YF, Zhang HY, Castellanos FX, Milham MP, 2010. Toward discovery science of human brain function. *PNAS* 107, 4734–4739. [PubMed: 20176931]
- 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, 2012. Test-retest reliability of resting-state connectivity network characteristics using fMRI and graph theoretical measures. *Neuroimage* 59, 1404–1412. [PubMed: 21888983]
- Cao H, Plichta MM, Schafer A, Haddad L, Grimm O, Schneider M, Esslinger C, Kirsch P, Meyer-Lindenberg A, Tost H, 2014. Test-retest reliability of fMRI-based graph theoretical properties during working memory, emotion processing, and resting state. *Neuroimage* 84, 888–900. [PubMed: 24055506]
- Chen G, Taylor PA, Haller SP, Kircanski K, Stoddard J, Pine DS, Leibenluft E, Brotman MA, Cox RW, 2018. Intra-class correlation: Improved modeling approaches and applications for neuroimaging. *Hum. Brain Mapp.* 39, 1187–1206. [PubMed: 29218829]
- Choe AS, Nebel MB, Barber AD, Cohen JR, Xu YT, Pekar JJ, Caffo B, Lindquist MA, 2017. Comparing test-retest reliability of dynamic functional connectivity methods. *Neuroimage* 158, 155–175. [PubMed: 28687517]
- Ciric R, Wolf DH, Power JD, Roalf DR, Baum GL, Ruparel K, Shinohara RT, Elliott MA, Eickhoff SB, Davatzikos C, Gur RC, Gur RE, Bassett DS, Satterthwaite TD, 2017. Benchmarking of participant-level confound regression strategies for the control of motion artifact in studies of functional connectivity. *Neuroimage* 154, 174–187. [PubMed: 28302591]
- Demetriou L, Kowalczyk OS, Tyson G, Bello T, Newbould RD, Wall MB, 2018. A comprehensive evaluation of increasing temporal resolution with multiband-accelerated protocols and effects on statistical outcome measures in fMRI. *Neuroimage* 176, 404–416. [PubMed: 29738911]

- Duarte CS, Monk C, Weissman MM, Posner J, 2020. Intergenerational psychiatry: a new look at a powerful perspective. *World Psychiatry* 19, 175–176. [PubMed: 32394546]
- Eyre M, Fitzgibbon SP, Ciarrusta J, Cordero-Grande L, Price AN, Poppe T, Schuh A, Hughes E, O’Keeffe C, Brandon J, Cromb D, Vecchiato K, Andersson J, Duff EP, Counsell SJ, Smith SM, Rueckert D, Hajnal JV, Arichi T, O’Muircheartaigh J, Batalle D, Edwards AD, 2021. The Developing Human Connectome Project: typical and disrupted perinatal functional connectivity. *Brain*.
- Finn ES, Shen X, Scheinost D, Rosenberg MD, Huang J, Chun MM, Papademetris X, Constable RT, 2015. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nat. Neurosci.* 18, 1664–1671. [PubMed: 26457551]
- Gao W, Alcauter S, Elton A, Hernandez-Castillo CR, Smith JK, Ramirez J, Lin W, 2015. Functional network development during the first year: relative sequence and socioeconomic correlations. *Cereb. Cortex* 25, 2919–2928. [PubMed: 24812084]
- Glasser MF, Coalson TS, Bijsterbosch JD, Harrison SJ, Harms MP, Anticevic A, Van Essen DC, Smith SM, 2018. Using temporal ICA to selectively remove global noise while preserving global signal in functional MRI data. *Neuroimage* 181, 692–717. [PubMed: 29753843]
- Gordon EM, Laumann TO, Gilmore AW, Newbold DJ, Greene DJ, Berg JJ, Ortega M, Hoyt-Drazen C, Gratton C, Sun H, 2017. Precision functional mapping of individual human brains. *Neuron* 95, 791–807 e797. [PubMed: 28757305]
- Gratton C, Kraus BT, Greene DJ, Gordon EM, Laumann TO, Nelson SM, Dosenbach NU, Petersen SE, 2019. Defining individual-specific functional neuroanatomy for precision psychiatry. *Biol. Psychiatry* 88, 28–39. [PubMed: 31916942]
- Grayson DS, Fair DA, 2017. Development of large-scale functional networks from birth to adulthood: a guide to the neuroimaging literature. *Neuroimage* 160, 15–31. [PubMed: 28161313]
- Greve DN, Fischl B, 2009. Accurate and robust brain image alignment using boundary-based registration. *Neuroimage* 48, 63–72. [PubMed: 19573611]
- Guo CC, Kurth F, Zhou J, Mayer EA, Eickhoff SB, Kramer JH, Seeley WW, 2012. One-year test-retest reliability of intrinsic connectivity network fMRI in older adults. *Neuroimage* 61, 1471–1483. [PubMed: 22446491]
- Hashempour N, Tuulari JJ, Merisaari H, Lidauer K, Luukkonen I, Saunavaara J, Parkkola R, Lahdesmaki T, Lehtola SJ, Keskinen M, Lewis JD, Scheinin NM, Karlsson L, Karlsson H, 2019. A novel approach for manual segmentation of the Amygdala and Hippocampus in Neonate MRI. *Front Neurosci.* 13, 1025. [PubMed: 31616245]
- Horowitz SG, Braun AR, Carr WS, Picchioni D, Balkin TJ, Fukunaga M, Duyn JH, 2009. Decoupling of the brain’s default mode network during deep sleep. *Proc. Natl. Acad. Sci. U. S. A.* 106, 11376–11381. [PubMed: 19549821]
- Jenkinson M, Bannister P, Brady M, Smith S, 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* 17, 825–841. [PubMed: 12377157]
- Jovicich J, Minati L, Marizzoni M, Marchitelli R, Sala-Llonch R, Bartres-Faz D, Arnold J, Benninghoff J, Fiedler U, Roccatagliata L, Picco A, Nobili F, Blin O, Bombois S, Lopes R, Bordet R, Sein J, Ranjeva JP, Didic M, Gros-Dagnac H, Payoux P, Zoccatelli G, Alessandrini F, Beltramello A, Bargallo N, Ferretti A, Caulo M, Aiello M, Cavaliere C, Soricelli A, Parnetti L, Tarducci R, Floridi P, Tsolaki M, Constantinidis M, Drevelegas A, Rossini PM, Marra C, Schonknecht P, Hensch T, Hoffmann KT, Kuijper JP, Visser PJ, Barkhof F, Frisoni GB, Consortium P, 2016. Longitudinal reproducibility of default-mode network connectivity in healthy elderly participants: a multicentric resting-state fMRI study. *Neuroimage* 124, 442–454. [PubMed: 26163799]
- Kozberg M, Hillman E, 2016. Neurovascular coupling and energy metabolism in the developing brain. In: *Progress Brain Research*. Elsevier, pp. 213–242.
- Lee CW, Blanco B, Dempsey L, Chalia M, Hebden JC, Caballero-Gaudes C, Austin T, Cooper RJ, 2020. Sleep state modulates resting-state functional connectivity in neonates. *Front Neurosci.* 14, 347. [PubMed: 32362811]

- Linke AC, Wild C, Zubiaurre-Elorza L, Herzmann C, Duffy H, Han VK, Lee DSC, Cusack R, 2018. Disruption to functional networks in neonates with perinatal brain injury predicts motor skills at 8 months. *Neuroimage Clin.* 18, 399–406. [PubMed: 29487797]
- Lugo-Candelas C, Cha J, Hong S, Bastidas V, Weissman M, Fifer WP, Myers M, Talati A, Bansal R, Peterson BS, Monk C, Gingrich JA, Posner J, 2018. Associations between brain structure and connectivity in infants and exposure to selective serotonin reuptake inhibitors during pregnancy. *JAMA Pediatr.* 172, 525–533. [PubMed: 29630692]
- Makropoulos A, Robinson EC, Schuh A, Wright R, Fitzgibbon S, Bozek J, Counsell SJ, Steinweg J, Vecchiato K, Passerat-Palmbach J, Lenz G, Mortari F, Tenev T, Duff EP, Bastiani M, Cordero-Grande L, Hughes E, Tusor N, Tournier JD, Hutter J, Price AN, Teixeira R, Murgasova M, Victor S, Kelly C, Rutherford MA, Smith SM, Edwards AD, Hajnal JV, Jenkinson M, Rueckert D, 2018. The developing human connectome project: a minimal processing pipeline for neonatal cortical surface reconstruction. *Neuroimage* 173, 88–112. [PubMed: 29409960]
- Marchitelli R, Minati L, Marizzoni M, Bosch B, Bartres-Faz D, Muller BW, Wiltfang J, Fiedler U, Roccatagliata L, Picco A, Nobili F, Blin O, Bombois S, Lopes R, Bordet R, Sein J, Ranjeva JP, Didic M, Gros-Dagnac H, Payoux P, Zoccatelli G, Alessandrini F, Beltramello A, Bargallo N, Ferretti A, Caulo M, Aiello M, Cavaliere C, Soricelli A, Parnetti L, Tarducci R, Floridi P, Tsolaki M, Constantinidis M, Drevelegas A, Rossini PM, Marra C, Schonknecht P, Hensch T, Hoffmann KT, Kuijter JP, Visser PJ, Barkhof F, Frisoni GB, Jovicich J, Consortium P, 2016. Test-retest reliability of the default mode network in a multi-centric fMRI study of healthy elderly: effects of data-driven physiological noise correction techniques. *Hum. Brain Mapp.* 37, 2114–2132. [PubMed: 26990928]
- Mejia AF, Nebel MB, Barber AD, Choe AS, Pekar JJ, Caffo BS, Lindquist MA, 2018. Improved estimation of subject-level functional connectivity using full and partial correlation with empirical Bayes shrinkage. *Neuroimage* 172, 478–491. [PubMed: 29391241]
- Mejia AF, Nebel MB, Shou H, Crainiceanu CM, Pekar JJ, Mostofsky S, Caffo B, Lindquist MA, 2015. Improving reliability of subject-level resting-state fMRI parcellation with shrinkage estimators. *Neuroimage* 112, 14–29. [PubMed: 25731998]
- Nickerson LD, Smith SM, Ongur D, Beckmann CF, 2017. Using dual regression to investigate network shape and amplitude in functional connectivity analyses. *Front Neurosci.* 11, 115. [PubMed: 28348512]
- Noble S, Scheinost D, Constable RT, 2019. A decade of test-retest reliability of functional connectivity: a systematic review and meta-analysis. *Neuroimage* 203, 116157. [PubMed: 31494250]
- Noble S, Scheinost D, Finn ES, Shen XL, Papademetris X, McEwen SC, Bearden CE, Addington J, Goodyear B, Cadenhead KS, Mirzakhani H, Cornblatt BA, Olvet DM, Mathalon DH, McGlashan TH, Perkins DO, Belger A, Seidman LJ, Thermenos H, Tsuang MT, van Erp TGM, Walker EF, Hamann S, Woods SW, Cannon TD, Constable RT, 2017. Multisite reliability of MR-based functional connectivity. *Neuroimage* 146, 959–970. [PubMed: 27746386]
- Panman JL, To YY, van der Ende EL, Poos JM, Jiskoot LC, Meeter LHH, Dopfer EGP, Bouts M, van Osch MJP, Rombouts S, van Swieten JC, van der Grond J, Papma JM, Hafkemeijer A, 2019. Bias introduced by multiple head coils in MRI research: an 8 channel and 32 channel coil comparison. *Front. Neurosci.* 13, 729. [PubMed: 31379483]
- Pannunzi M, Hindriks R, Bettinardi RG, Wenger E, Lisofsky N, Martensson J, Butler O, Filevich E, Becker M, Lochstet M, Lindenberger U, Kuhn S, Deco G, 2018. Resting-state fMRI correlations: From link-wise unreliability to whole brain stability (vol 157, pg 250, 2017). *Neuroimage* 174, 599–604. [PubMed: 29289426]
- Parkes L, Fulcher B, Yucel 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. [PubMed: 29278773]
- Posner J, Cha J, Roy AK, Peterson BS, Bansal R, Gustafsson HC, Raffanella E, Gingrich J, Monk C, 2016. Alterations in amygdala-prefrontal circuits in infants exposed to prenatal maternal depression. *Transl. Psychiatry* 6, e935. [PubMed: 27801896]

- Power JD, Cohen AL, Nelson SM, Wig GS, Barnes KA, Church JA, Vogel AC, Laumann TO, Miezin FM, Schlaggar BL, Petersen SE, 2011. Functional network organization of the human brain. *Neuron* 72, 665–678. [PubMed: 22099467]
- Raz G, Saxe R, 2020. Learning in infancy is active, endogenously motivated, and depends on the prefrontal cortices. *Ann. Rev. Dev. Psychol.* 2, 247–268.
- Rudolph MD, Graham AM, Feczko E, Miranda-Dominguez O, Rasmussen JM, Nardos R, Entringer S, Wadhwa PD, Buss C, Fair DA, 2018. Maternal IL-6 during pregnancy can be estimated from newborn brain connectivity and predicts future working memory in offspring. *Nat. Neurosci.* 21, 765–772. [PubMed: 29632361]
- Salehi M, Greene AS, Karbasi A, Shen XL, Scheinost D, Constable RT, 2020. There is no single functional atlas even for a single individual: Functional parcel definitions change with task. *Neuroimage* 208, 116366. [PubMed: 31740342]
- Shi F, Yap PT, Wu G, Jia H, Gilmore JH, Lin W, Shen D, 2011. Infant brain atlases from neonates to 1- and 2-year-olds. *PLoS One* 6, e18746. [PubMed: 21533194]
- Shirer WR, Jiang H, Price CM, Ng B, Greicius MD, 2015. Optimization of rs-fMRI pre-processing for enhanced signal-noise separation, test-retest reliability, and group discrimination. *Neuroimage* 117, 67–79. [PubMed: 25987368]
- Shrout PE, Fleiss JL, 1979. Intraclass Correlations - Uses in Assessing Rater Reliability. *Psychol. Bull.* 86, 420–428. [PubMed: 18839484]
- Smith SM, 2002. Fast robust automated brain extraction. *Hum. Brain Mapp.* 17, 143–155. [PubMed: 12391568]
- Smith SM, Brady JM, 1997. SUSAN - A new approach to low level image processing. *Int. J. Comput. Vision* 23, 45–78.
- Smith SM, Fox PT, Miller KL, Glahn DC, Fox PM, Mackay CE, Filippini N, Watkins KE, Toro R, Laird AR, Beckmann CF, 2009. Correspondence of the brain's functional architecture during activation and rest. *Proc. Natl. Acad. Sci. U. S. A.* 106, 13040–13045. [PubMed: 19620724]
- Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TEJ, Johansen-Berg H, Bannister PR, De Luca M, Drobnjak I, Flitney DE, Niazy RK, Saunders J, Vickers J, Zhang YY, De Stefano N, Brady JM, Matthews PM, 2004. Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* 23, S208–S219. [PubMed: 15501092]
- Somandepalli K, Kelly C, Reiss PT, Zuo XN, Craddock RC, Yan CG, Petkova E, Castellanos FX, Milham MP, Di Martino A, 2015. Short-term test-retest reliability of resting state fMRI metrics in children with and without attention-deficit/hyperactivity disorder. *Dev. Cogn. Neurosci.* 15, 83–93. [PubMed: 26365788]
- Termenon M, Jaillard A, Delon-Martin C, Achard S, 2016. Reliability of graph analysis of resting state fMRI using test-retest dataset from the Human Connectome Project. *Neuroimage* 142, 172–187. [PubMed: 27282475]
- Wang Y, Bernanke J, Peterson BS, McGrath P, Stewart J, Chen Y, Lee S, Wall M, Bastidas V, Hong S, Rutherford BR, Hellerstein DJ, Posner J, 2019. The association between antidepressant treatment and brain connectivity in two double-blind, placebo-controlled clinical trials: a treatment mechanism study. *Lancet Psychiatry* 6, 667–674. [PubMed: 31248841]
- Wiggins GC, Triantafyllou C, Potthast A, Reykowschi A, Nittka M, Wald LL, 2006. 32-channel 3 Tesla receive-only phased-array head coil with soccer-ball element geometry. *Magn. Reson. Med.* 56 (1), 216–223. [PubMed: 16767762]
- Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, Gerig G, 2006. User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *Neuroimage* 31, 1116–1128. [PubMed: 16545965]

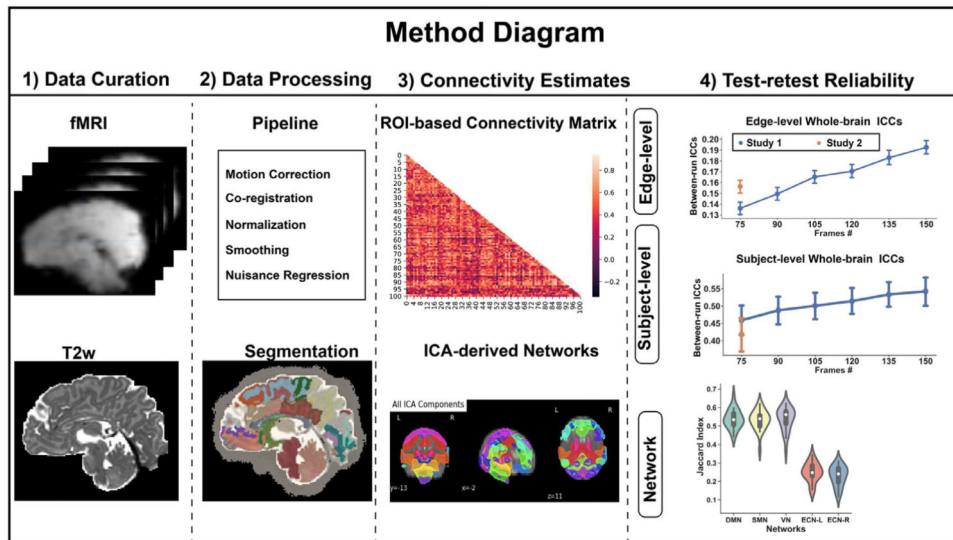


Fig. 1.
 Data Analytic Plan in this study.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

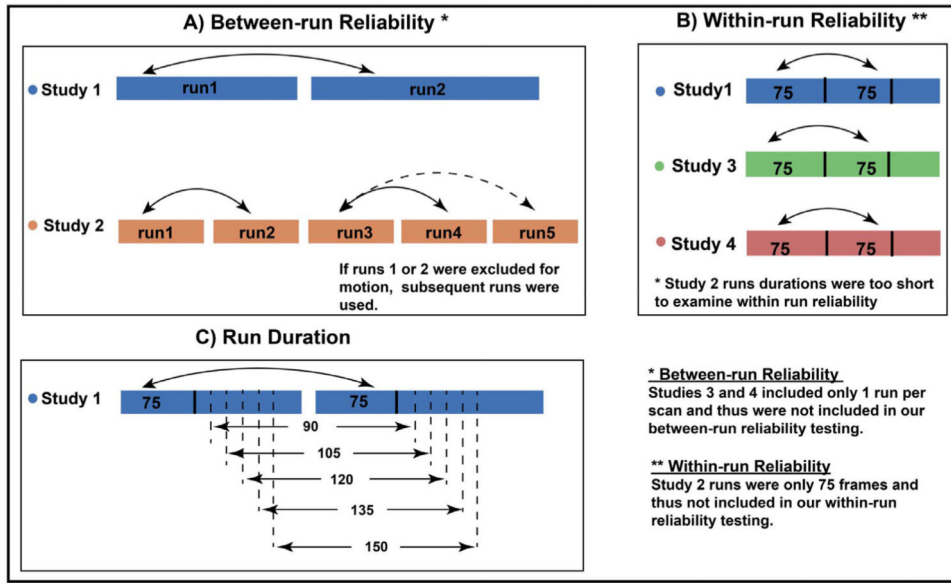


Fig. 2. Schematics to estimate **A)** between-run test-retest reliability estimates, **B)** within-run reliability across study sites, **C)** the effect of run duration on between-run reliability.

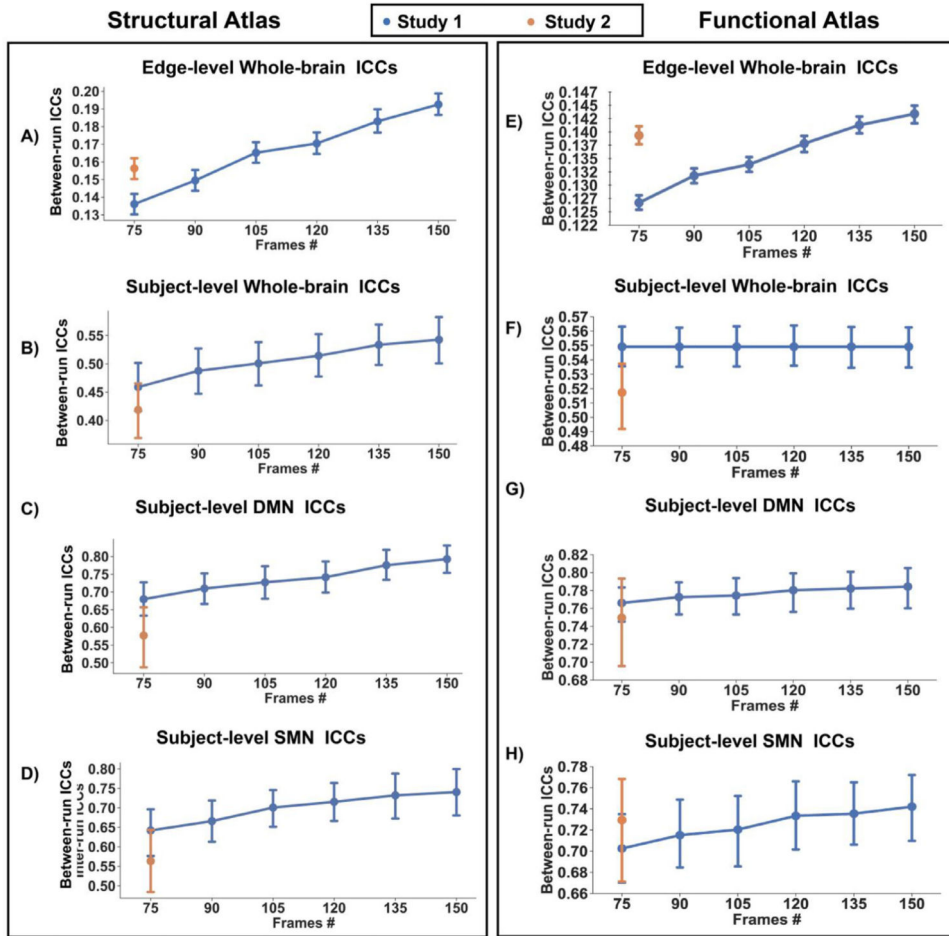


Fig. 3. Test-retest reliability defined with a structural atlas approaches: **A)** edge level, and subject level for **B)** whole brain, **C)** DMN, and **D)** SMN
 Reliability defined with a functional atlas approach: **E)** edge level, and subject level for **F)** whole brain, **G)** DMN, and **H)** SMN

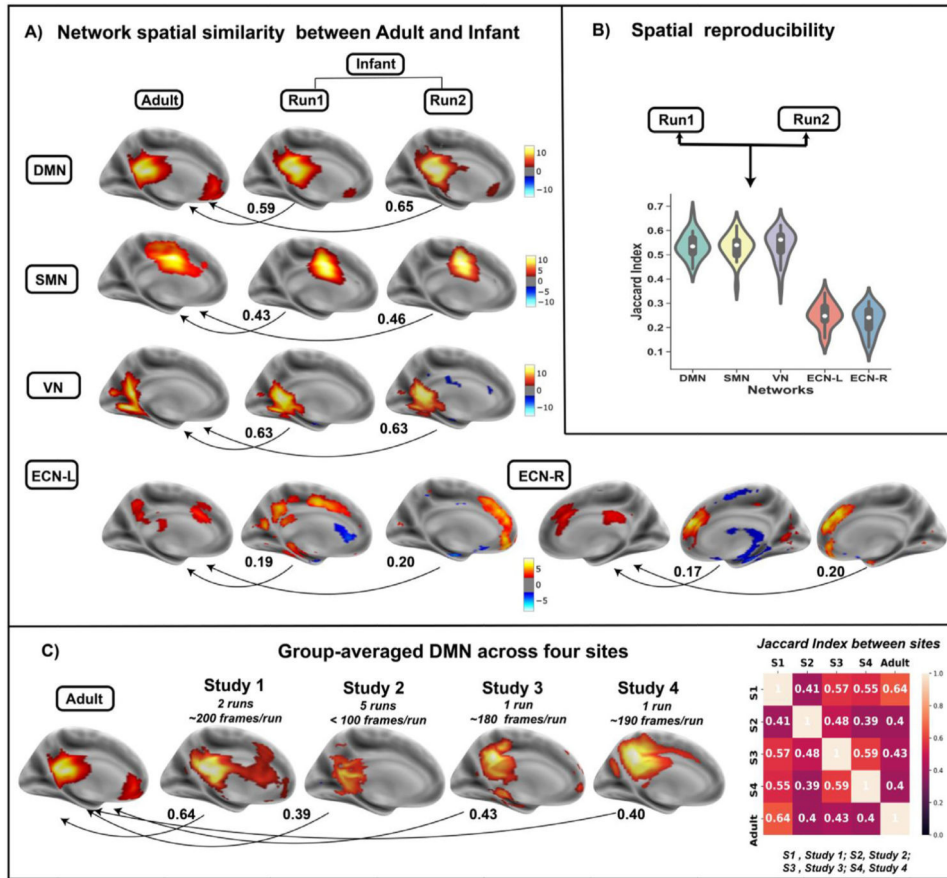


Fig. 4. A) Spatial similarity(measured by Jaccard Index) between five ICA-derived networks (DMN, SMN, VN, ECN-L and ECN-R) in infants and corresponding ones in adults (Smith 2009) across runs. B) Spatial reproducibility of five networks across two runs in study 1. C) Group-averaged DMN for four study sites as well as their pairwise similarities with DMN in adults.

Table 1

Demographics and MRI parameters by study.

| | Missing ^a | Study 1 | Study 2 | Study 3 | Study 4 | p-value |
|--|----------------------|------------------------|------------------------|------------------------|------------------------|---------|
| Number of scans | | 43 | 23 | 27 | 26 | |
| PMA, mean (SD) | 9 | 46.70 (4.63) | 42.30 (1.20) | 42.04 (0.90) | 49.53 (4.512) | <0.001 |
| Sex, n (%) | F 8 M | 19 (51.4) 18 (48.6) | 12 (54.5) 10 (45.5) | 14 (53.8) 12 (46.2) | 15 (57.7) 11 (42.3) | 0.991 |
| Scanner | | GE 3T | GE 3T | Phillips 3T | GE 3T | |
| TR (sec) | | 2.0 | 2.2 | 2.3 | 2.0 | |
| Runs | | 2 | 5 | 1 | 1 | |
| Numb. frames/ run | | 210 | 102 | 210 | 210 | |
| Number of usable frames/run after scrubbingmean [min, max] | | 200[185, 216] | 90[76,102] | 179[126,210] | 190[153,210] | |

PMA, postmenstrual age; M, male, F, female

Study 1: Environmental influences on child health outcomes – Boricua Youth Study (ECHO-BYS) at the New York State Psychiatric Institute (NYSPI); Study 2: Serotonin system and neurodevelopment; Study 3, Brazil Babies Study; Study 4, ECHO-BYS at the University of Puerto Rico

^aMissing PMA and/or sex occurred in Study 1 and 3.