Inter-Scanner Reliability of Graph-Theoretic Brain Network Metrics

Thomas Welton[1], Dorothee P Auer[1], and Robert A Dineen[1]
[1]*Sir Peter Mansfield Imaging Centre, School of Medicine, University of Nottingham, Nottingham, Nottinghamshire, United Kingdom*

Target Audience: Those interested in applying graph-theory to brain MRI data.

Purpose: Multiple studies have examined the reproducibility of graph metrics of functional brain networks in the same MRI scanner over time[1] (intra-rater reliability), but to our knowledge no studies to date have investigated reproducibility of these measures between scanners (inter-rater reliability). This is a crucial prerequisite to the wider application of graph metrics in clinical neuroscience and, in particular, to future adoption in multicentre treatment trials. We therefore aimed to test the inter-scanner reliability of graph-theoretic brain network metrics.

Methods: Data were downloaded from the fBIRN repository (accession number 2007-BDR-6UHZ1). We used resting-state fMRI scans from five healthy subjects (median age 25, range 20-28, all right-handed males) acquired on 10 different scanners (sequence: EPI or spiral GRE, plane: axial AC-PC, TR: 3000 ms, TE: 30 ms for 3T and 4T or 40ms for 1.5T, FA: 90°, voxel size: 3.475×3.475×4.000 mm contiguous, matrix size: 64×64×35 voxels in 1 shot, FOV: 22.24 cm, BW: ≥ ± 100 kHz, slice acquisition: interleaved, scan time: 4.5 minutes). Data were pre-processed using FSL: motion correction, slice timing correction, brain extraction, spatial smoothing, highpass temporal filtering and registration to a standard space were performed. BOLD timeseries were extracted from 48 cortical and 14 subcortical regions of interest using the Harvard-Oxford structural atlas. Pearson correlation coefficients were calculated for each pair of regions to form weighted functional connectivity association matrices. These were thresholded at a range of densities (0.1 – 0.4 in steps of 0.01). At each density, the following metrics were calculated: clustering coefficient, characteristic path length, small-worldness, global efficiency and modularity. The area under the curve was calculated for each metric across densities, to give a single summary value. Inter-scanner and inter-subject reliability was quantified using the coefficient of variation (CV) and intra-class correlation coefficients (ICC) for agreement and consistency. ICCs are interpreted as follows: <0.40, "poor"; 0.40-0.59, "fair"; 0.60-0.74, "good"; >0.74, "excellent"[2]. We also tested the relationship between scanner field strength and reliability, and whether using graph metrics based on an average of the 4 resting-state fMRI scans in each visit would improve reliability.

Results: Most ICCs were graded as "poor", and only the consistency ICCs for clustering and small-worldness were graded as "fair". Tests for equality of ICCs based on subsets of scanners, classified into either 1.5T or 3.0/4.0T, showed neither group to be significantly more reliable. However, CVs based on an average of 4 scans were significantly lower than those based on a single scan, for both inter- and intra-scanner tests (one-tailed sign tests: $p < 0.0001$ and $p < 0.0001$, respectively).

Discussion: Reliability of graph-theoretic functional brain networks between scanners was generally poor. Our results were in accordance with a previous study in the same dataset testing the reliability of contrast-to-noise and percent signal change measurements[3]. In contrast to our finding, reliability between subjects within the same scanner can often be "good" or "excellent"[1], which suggests that the variability seen here is inherent in the differences between scanners, rather than in true biological variation over time. Other possible limiting factors for reliability were: quality and duration of scan, subject motion, graph density threshold or region of interest size. Averaging graph-theoretic metrics from multiple resting state fMRI acquisitions in each scanner provides one means to improve reliability between different scanners.

Conclusion: **Graph metrics derived from human brain fMRI data are not yet reliable enough between scanners for use as surrogate outcomes in multisite clinical trials.**

References: [1] Andreotti et al, 2014. Repeatability analysis of global and local metrics of brain structural networks. *Brain Connectivity* 4(3):203-20. [2] Shrout & Fleiss, 1979. Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin* 86(2):420. [3] Friedman et al, 2008. Test–retest and between-site reliability in a multicenter fMRI study. *Human Brain Mapping* 29(8):958-972.

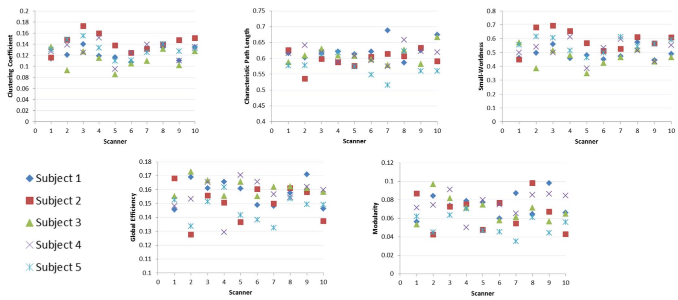

Figure 1. Dot plots of inter-scanner reliability of graph metrics, based on a single fMRI scan. From left to right: (top row) clustering coefficient, characteristic path length, small-worldness, (bottom row) global efficiency and modularity.

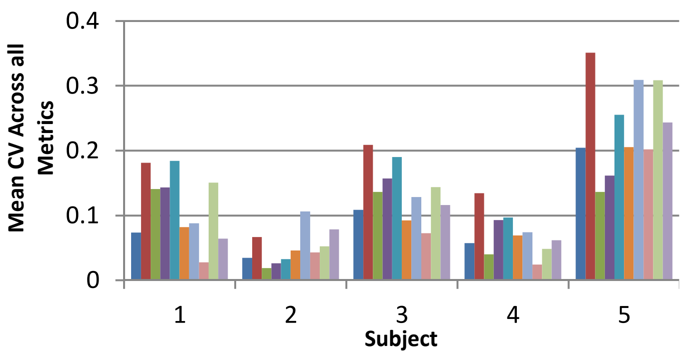

Figure 2. Line graph showing, for each subject, the mean CV (across all graph metrics) for each scanner. The coloured bars represent the 10 scanners.