A Framework for Analysis of Living Phantom Data in a Multicenter DTI study

L. Walker[1], N. Lange[2], L-C. Chang[3], C. Pierpaoli[1], and .. the Brain Development Cooperative Group[4]

[1]STBB, NICHD, National Institutes of Health, Bethesda, MD, United States, [2]Departments of Psychiatry and Biostatistics, Harvard Schools of Medicine and Public Health, Boston, MA, United States, [3]Department of Electronic Engineering adn Computer Science, The Catholic University of America, Washington, DC, United States, [4]www.NIH-PediatricMRI.org

Introduction: The NIH MRI Study of Normal Brain Development (www.NIH-PediatricMRI.org) is a comprehensive study of human brain development comprising neuroimaging data from ~500 typically developing children between the ages of 7 days and 18 years. All imaging data in this study will be publicly available. While multicenter studies have a great advantage in allowing for increased numbers of participants [1], the variability of acquired data is greater than in single-center studies, due to the added confound of inter-site variability [2]. As multicenter DTI studies gain in popularity, it is important to assess the impact of these sources of variability in order to meaningfully pool quantitative imaging data across sites. Thus, a living phantom was scanned at multiple time points during the project as a control. In this work we propose tools for the assessment of living phantom data in multicenter DTI studies, and provide results using these tools on the living phantom DTI imaging data from the NIH MRI Study of Normal Brain Development.

Methods: The living phantom is a healthy adult male aged 51 years at first scan. Twenty-two successful DTI scans were acquired at 5 sites over 4 years. It is assumed that the living phantom brain is stable over this time and any aging related variability is much smaller than inter- and intra-site variability. DTI data was acquired at 1.5T (GE or Siemens) with 6 non-collinear directions at b=1000s/mm^2 and 1 b=0s/mm^2, repeated 4 times for 28 brain volumes, no averaging. Living phantom datasets were assessed for gross artifacts. Two were rejected due to severe ghosting and low SNR. The remaining 20 datasets were pre-processed using TORTOISE (www.tortoisedti.org)[3] to correct for motion and eddy distortions including appropriate rotations of the b-matrix [4] and to register all datasets into a common space. Data were processed again for EPI distortion correction [5].

We propose two tools that are ideally suited to a living phantom situation where the individual time points should be morphologically identical, allowing for meaningful voxel-wise measurements of variance. 1) Inspection of outlier datasets based on the voxel-wise median of tensor-derived metrics. Assuming that acquisitions are balanced across sites, and less than 50% of datasets are corrupted by artifacts, the median will be a measure of the central tendency for the samples. Subtracting each dataset's fractional anisotropy (FA) and trace of the diffusion tensor (TR) images from the median image will allow identification of corrupted data with significant deviation from median (**Fig 1**). Images identified as outliers are inappropriate for a parametric analysis of inter- and intra-site variability, thus are corrected if possible, or removed. 2) Comparison of between-site and within-site variance to assess the contributions of each of these sources to overall variance. Similar to a traditional ANOVA, one can compute the within-site (*intra-site*) variability, defined as the square root of the mean of the scan variances from site to site, and the between-site (*inter-site*) variability, defined as the standard deviation of the mean scans across sites. These measures can be computed at each step of a processing pipeline to assess if corrections affect the inter- and intra-site variability.

Results: Initial median analysis revealed two outliers showing large deviation with elevated TR (**Fig 1**) and reduced FA. Further investigation suggested that these data were acquired with an incorrect maximum b-value of ~1500 rather than 1000s/mm^2. This was corrected by adjusting the b-matrix. Isolated outliers may not be identified in parametric variance testing, highlighting the importance of this outlier detection step.

Figure 1: TR for each dataset subtracted from the median TR image (intensity range -1500 – 1500μm^2/s). All datasets show some minor deviations from the median. Red circles indicate outlier data points requiring correction or removal.

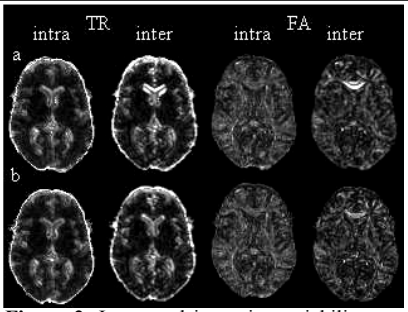

Figure 2: Inter- and intra-site variability at different steps in processing pipeline. FA range 0 – 0.3, TR range 0 – 2000μm^2/s

Variance analysis was performed on FA and TR (**Fig 2**) at different stages of processing a) before, and b) after EPI correction. In the ideal case, the inter- and intra-site variability should be of the same magnitude indicating no differences between sites; this is the basis of the ANOVA F-test under a null hypothesis of no site effects. In **Fig 2** the inter-site variability is slightly elevated compared to the intra-site variability overall in the parenchyma, suggesting increased variance in the data due to site differences. The stronger effect, however, is higher variability at CSF-tissue interfaces, likely due to inconsistent brain morphology. This effect is also apparent in **Fig 1**. Large inter-site variability is seen in the frontal lobes and genu of the corpus callosum; brain regions affected by susceptibility-induced EPI distortions. Variance analysis after EPI correction shows that this regionally increased inter-site variability is dramatically reduced.

Discussion and Conclusion: Use of a living phantom in multicenter studies is a valuable tool for assessing potential issues with data quality, and understanding the contributions to variance in pooled data. The combination of the two proposed tools allows for identification of outlier datasets, and investigation of sources of inter- and intra-site variability. Most of the regions with high variability are at tissue interfaces, suggesting that inconsistent morphology is a significant issue. This is curious, considering we are using a single subject. In this study, differences in phase encode direction between scanners (Anterior-Posterior with either a positive or negative direction) resulted in significant morphological differences between sites. EPI correction is able to correct the large regional variability, but is not able to account for the more global tissue interface effects. In designing a multicenter DTI study, it is key that only a single subject is used as a living phantom in order to avoid this major confound of inter-subject anatomical variability, which would mask the more subtle and important effects of equipment-related variability.

References:[1] Vollmar, C, et al, NeuroImage,2001, 51 1384-1394 [2] Pfefferbaum, A, et al, JMRI, 2003,18(4),749-761 [3] Pierpaoli, C, et al, ISMRM 18th Ann. Mtg, Stockholm, #1597 [4] Rohde, GK, et al, MRM 2004. 51(1), p 103-14 [5] Wu, M, et al, MICCAI 2008 11(Pt 2): p321-9