A Novel Method for Determining the Reliability of Diffusion-Weighted Imaging Data

Karl G Helmer[1], Ming-Chung Chou[2], Ronny I Preciado[3], Allen Song[4], Jessica Turner[5], Barjor Gimi[6], and Susumu Mori[7,8]

[1]Radiology, Massachusetts General Hospital, Charlestown, MA, United States, [2]Department of Medical Imaging and Radiological Sciences, Kaohsiung Medical University, Kaohsiung, Taiwan, Kaohsiung, Taiwan, [3]Massachusetts General Hospital, Charlestown, MA, United States, [4]Radiology, Duke University, Durham, NC, United States, [5]Translational Neuroscience, The Mind Research Network, Albuquerque, NM, United States, [6]Radiology and Medicine, Dartmouth-Hitchcock Medical Center, Lebanon, NH, United States, [7]Radiology, Johns Hopkins School of Medicine, Baltimore, MD, United States, [8]Kennedy Kreiger Institute, Baltimore, MD, United States

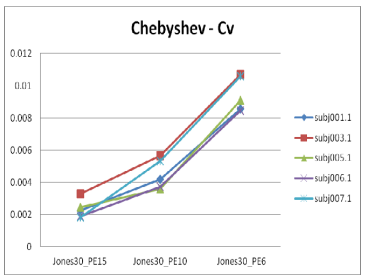
Target Audience: Researchers who seek to a quantitative method to characterize the reliability of human brain diffusion-weighted imaging (DWI) performance for a single site or group of sites.

Purpose: The goal of this work is to develop a method to characterize and compare the DWI performance of one or more sites using commonly calculated DWI metrics. Because multi-site studies have become ubiquitous, a method is needed to quantitatively track the quality of DW data through soft- and hard-ware changes. Studies have been undertaken to look at diffusion tensor imaging (DTI) metric reproducibility, e.g. [1], but little work has been done that includes a full complement of commonly varied protocol parameters and site effects. In this work, we investigated 14 different metrics from several different families of histogram-distance metrics to characterize the effects of parameters such as number of gradient directions, data quality, echo time and field strength as well as vendor and site effects. These metrics vary in computational complexity and were investigated to identify a smaller group a metrics that could be used to adequately characterize histograms of fractional anisotropy and mean diffusivity from human brain DWI.

Methods: Five locally-recruited subjects were scanned at each of five sites. Three scanner vendors were represented: Siemens (2 sites), GE (1 site), and Philips (2 sites). Five of the sites had field strengths of 3.0T; the 6th site (Philips2) was 1.5T. Ten DTI scans were performed on each subject, using the Jones 30 set of diffusion-weighted directions (DWD) and 5 b=0 scans. Each Jones30/5 b=0 set was defined to be one "scan-time unit" (STU). Other protocol parameters include: b-value of 1000 s/mm^2, 2.5 mm^3 isotropic voxels, acquired matrix size: 96 x 96, full k-space coverage, FOV: 240 x 240 mm, number of slices: 25, parallel imaging: SENSE (p = 2) for Philips and GRAPPA for Siemens, 1 average, TR/TE (ms) were: Siemens = 4000/98.0 (site1), 3800/98.0 (site2); GE = 5200/99.5; Philips = 4000/101.19 (site2), 4000/100.00 (site1). In these scans TR/TE were harmonized as much as possible given the constraints set by different manufacturers. The final TE was dictated by the longest minimum TE achievable over all the sites. A separate set of 5 subjects was acquired on the GE scanner using their minimum TE (69.8ms), to determine the effect on FA of minimizing TE, a commonly used scenario. Data sets with different numbers of STU (1-10) were constructed by concatenating a sequentially increasing number of data sets before calculation of the tensors and associated metrics. The data were resampled to 15, 10, and 6 gradient directions and the FA and MD values recalculated.

Analysis: Whole-brain FA/MD histograms of brain voxel values were calculated using 0.01/0.00005-wide bins over the ranges [0.0,1.0]/[0.0,0.004 mm^2/s]. To assess within-site variability, the original Jones30 data and each resampling was compared to each other giving 6 metric values per subject. Means and standard deviations were then calculated for each subject and for each site as a whole for SNR = 1/10. For the between-site case, the Jones30 data was compared between subjects at different sites and each resampling using the SNR=1/10 data. To determine metric cutoff values for statistical significance we followed [2]: normal and log-normal histograms were simulated using random samples of elements from those distributions. Samples from the log-normal distribution were generated using the Metropolis-Hastings algorithm; the normal distribution samples were generated using the Box-Muller method with the addition of a Bays-Durham shuffle to remove low-order serial correlations. A total of 61245 histogram distances for each metric were calculated from 351 histograms by comparing each to all of the (unique) others. Using the resulting distribution of these distance metric values, the 95th percentile value was determined and used as a cutoff value, d_c. This process was repeated for 18/12 normal/log-normal-distribution variance values. In addition, 9 different number of histogram elements, N, in the range [10,000,90,000] were calculated. This gives a family of curves of d_c versus N with each curve having a specific distribution variance. Knowing the experimental variance for a given experiment the applicable d_c.value can then be used to determine the statistical significance of the calculated histogram distance, d, for the experimental data.

Results and Discussion: The Fig. shows an example of the within-site comparison data for one subject at a Siemens site for the Chebyshev distance metric. In this figure the histogram distance metric is plotted for each subject when the Jones30 FA histogram is compared to each of the other resamplings. This metric gave the largest variance of the metric values. As can be seen the metric value increases as the Jones30 "gold standard' data is compared to data with decreasing number of directions. When comparisons were made across data of different SNR values, the metrics increased as the SNR=10 data was compared with that for lower SNR data. Statistical significance was determined using the simulations described above. We have found that this method can distinguish between data of different vendors, field strength and TE values, while the results are generally consistent within sites.



References: [1] Pfefferbaum, et al., JMRI, 18:427 (2003), [2] Bernas et al., Cytometry Part A 2008;73A(8):715-726.