NeuroImage xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

The impact of quality assurance assessment on diffusion tensor imaging outcomes in a large-scale population-based cohort

David R. Roalf [a,*], Megan Quarmley [a], Mark A. Elliott [b], Theodore D. Satterthwaite [a], Simon N. Vandekar [a,e], Kosha Ruparel [a], Efstathios D. Gennatas [a], Monica E. Calkins [a], Tyler M. Moore [a], Ryan Hopson [a], Karthik Prabhakaran [a], Chad T. Jackson [a], Ragini Verma [b,c], Hakon Hakonarson [d], Ruben C. Gur [a,b], Raquel E. Gur [a,b]

[a] Neuropsychiatry Section, Department of Psychiatry, USA
[b] Department of Radiology, University of Pennsylvania, Perelman School of Medicine, USA
[c] Section of Biomedical Image Analysis, USA
[d] Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA, USA
[e] Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, PA 19104, USA

ARTICLE INFO

Article history:
Received 1 July 2015
Accepted 24 October 2015
Available online xxxx

Keywords:
Diffusion tensor imaging
Automated quality assurance
Motion
Adolescence
Brain maturation

ABSTRACT

Background: Diffusion tensor imaging (DTI) is applied in investigation of brain biomarkers for neurodevelopmental and neurodegenerative disorders. However, the quality of DTI measurements, like other neuroimaging techniques, is susceptible to several confounding factors (e.g., motion, eddy currents), which have only recently come under scrutiny. These confounds are especially relevant in adolescent samples where data quality may be compromised in ways that confound interpretation of maturation parameters. The current study aims to leverage DTI data from the Philadelphia Neurodevelopmental Cohort (PNC), a sample of 1601 youths with ages of 8–21 who underwent neuroimaging, to: 1) establish quality assurance (QA) metrics for the automatic identification of poor DTI image quality; 2) examine the performance of these QA measures in an external validation sample; 3) document the influence of data quality on developmental patterns of typical DTI metrics.

Methods: All diffusion-weighted images were acquired on the same scanner. Visual QA was performed on all subjects completing DTI; images were manually categorized as Poor, Good, or Excellent. Four image quality metrics were automatically computed and used to predict manual QA status: Mean voxel intensity outlier count (MEANVOX), Maximum voxel intensity outlier count (MAXVOX), mean relative motion (MOTION) and temporal signal-to-noise ratio (TSNR). Classification accuracy for each metric was calculated as the area under the receiver-operating characteristic curve (AUC). A threshold was generated for each measure that best differentiated visual QA status and applied in a validation sample. The effects of data quality on sensitivity to expected age effects in this developmental sample were then investigated using the traditional MRI diffusion metrics: fractional anisotropy (FA) and mean diffusivity (MD). Finally, our method of QA is compared with DTIPrep.

Results: TSNR (AUC = 0.94) best differentiated Poor data from Good and Excellent data. MAXVOX (AUC = 0.88) best differentiated Good from Excellent DTI data. At the optimal threshold, 88% of Poor data and 91% Good/Excellent data were correctly identified. Use of these thresholds on a validation dataset (n = 374) indicated high accuracy. In the validation sample 83% of Poor data and 94% of Excellent data was identified using thresholds derived from the training sample. Both FA and MD were affected by the inclusion of poor data in an analysis of age, sex and race in a matched comparison sample. In addition, we show that the inclusion of poor data results in significant attenuation of the correlation between diffusion metrics (FA and MD) and age during a critical neurodevelopmental period. We find higher correspondence between our QA method and DTIPrep for Poor data, but we find our method to be more robust for apparently high-quality images.

Conclusion: Automated QA of DTI can facilitate large-scale, high-throughput quality assurance by reliably identifying both scanner and subject induced imaging artifacts. The results present a practical example of the confounding effects of artifacts on DTI analysis in a large population-based sample, and suggest that estimates of data quality should not only be reported but also accounted for in data analysis, especially in studies of development.

* Corresponding author at: Department of Psychiatry, Neuropsychiatry Section, Brain Behavior Laboratory, 10th Floor, Gates Building, Hospital of the University of Pennsylvania, Philadelphia, PA 19104, USA.
E-mail address: roalf@upenn.edu (D.R. Roalf).

http://dx.doi.org/10.1016/j.neuroimage.2015.10.068
1053-8119/

Introduction

Diffusion tensor imaging (DTI) is an important magnetic resonance imaging (MRI) technique in the investigation and identification of brain biomarkers in typical neurodevelopment, cognitive aging and neuropsychiatric syndromes. DTI is based on the quantification of the random Brownian motion of protons, which enables the measurement of the spatial organization of brain tissue (Basser et al., 1994; Le Bihan et al., 2001). Specifically, DTI provides contrasts that are sensitive to intra-voxel white matter microstructure (Basser and Pajevic, 2000) and produces results that are consistent with the major white matter pathways detailed in animal models and in retinotopic studies in the human brain (Conturo et al., 1999; Le Bihan, 2003). An extensive DTI literature details findings in normal development (e.g., Ladouceur et al., 2012; Lenroot and Giedd, 2010; Oishi et al., 2013; Yoshida et al., 2013) and neuropsychiatric conditions, including schizophrenia (Roalf et al., 2015; Wheeler and Voineskos, 2014), autism (Konrad and Eickhoff, 2010; Travers et al., 2012) and Alzheimer's disease (Radanovic et al., 2013; Zhang et al., 2014). Moreover, large consortia, such as the Human Connectome Project (Van Essen et al., 2012) and the Alzheimer's Disease Neuroimaging Initiative (Jack et al., 2010) rely on DTI data as a major outcome measure. However, DTI is not without practical challenges that affect the reliability and reproducibility of results (Le Bihan et al., 2006).

Neuroimaging data confounds, in particular head motion, are quite pertinent in human samples (Liu et al., 2015; Power et al., 2012), especially children (Yoshida et al., 2013) and adolescents (Satterthwaite et al., 2012). For example, the confounding influence of head motion on resting-state functional connectivity has received substantial attention (Power et al., 2012; Satterthwaite et al., 2012; Van Dijk et al., 2012). Similar effects are evident in structural MRI (Reuter et al., 2015) and pediatric DTI samples (Lauzon et al., 2013; Yoshida et al., 2013). DTI measurements, in general, are reliable as they are insensitive to B_1 errors, however, DTI is strongly dependent on gradient calibration and errors in gradient amplitude, direction and linearity, can contribute to inaccurate measurement (Conturo et al., 1995). Nonetheless, the influence of data quality using DTI is understudied and often ignored. Beyond head motion, the quality of DTI measurements is susceptible to several confounds including eddy currents, scanner artifacts (e.g., noise spikes) and susceptibility artifacts (Anderson, 2001; Bastin et al., 1998; Skare et al., 2000), which have come under scrutiny (Heim et al., 2004; Jones et al., 1999, 2013; Lauzon et al., 2013; Owens et al., 2012; Tournier et al., 2011; Yendiki et al., 2014), and are the focus of several new methods seeking to mitigate their impact (Li et al., 2013, 2014; Oguz et al., 2014). These confounds likely contribute to inaccuracies in the tensor fitting of DTI data (Le Bihan et al., 2006). For example, head-motion was found to induce group differences between autistic and typically developing children in DTI (Yendiki et al., 2014). Most importantly, the use of head-motion as a nuisance regressor during statistical modeling reduced this effect. Yet, most developmental and clinical studies using DTI fail to report procedures for quality control and its impact on results. It is likely that unaccounted for artifacts result in suboptimal tensor estimation and thus may negatively influence commonly derived DTI metrics, such as fractional anisotropy, mean diffusivity, and estimates of tractography. In addition, because data quality is often systematically related to a phenotype of interest (e.g., age, diagnosis, cognition, symptom severity) and that data quality is inherently subject dependent (e.g., correlation between age and motion), low quality data has the potential to obscure the presence of real effects or produce spurious associations with study phenotypes.

Despite such dangers, automated measures for quality assurance (QA) of DTI data remain limited. Manual inspection of multivolume DTI data is time consuming, subjective and potentially susceptible to operator bias, and translates poorly to large-scale imaging studies. Studies of noise in DTI provide a useful framework for identifying how such noise affects diffusion properties (Ding et al., 2005; Farrell et al., 2007; Hasan, 2007; Skare et al., 2000). Several recent studies indicate promise for implementing automatically derived quality assurance metrics that reduce the amount of manual QA effort, including measures of signal-to-noise and the use of outlier detection, to quantify data quality prior to image processing (Lauzon et al., 2013; Li et al., 2013, 2014; Oguz et al., 2014). However, much of this work has used relatively small samples or simulated data, and none have focused primarily on a neurodevelopment sample (although Lauzon et al., 2013 present data in a large pediatric sample). Finally, there is lack of corroboration of derived metrics in a validation sample.

The overall goal of the current study is to determine which image QA metrics are most reliable in the automatic detection of poor DTI data. We manually evaluate over 1500 DTI data sets from the Philadelphia Neurodevelopmental Cohort (PNC; (Gur et al., 2014; Satterthwaite et al., 2014), and automatically derive QA measures. This approach will be useful in the current cross-sectional sample, in concurrent or longitudinal studies, and generalizable to most DTI studies. Importantly, all DTI data in the PNC was acquired within a 30-month period using the same MRI scanner, head-coil and DTI protocol. In addition, 25% of the sample returned for a follow-up DTI scan approximately two years later, thus providing a unique validation sample. Our goals are: 1) leverage DTI data from the PNC, a sample of 1601 youth between the age of 8–21 who underwent neuroimaging, to determine automated quality assurance metrics that will aid in the automatic identification of poor DTI image quality; 2) test these QA measures in a follow-up sample; and, 3) determine the influence of data quality on typical DTI metrics (e.g., FA and MD), 4) measure changes introduced by including poor data in the correlations between FA/MD and age and 5) compare our QA processes to a previous published DTI QA tool, DTIPrep (Oguz et al., 2014). As described below, results indicate automated QA of DTI can facilitate large-scale, high-throughput analysis by reliably identifying poor quality data and systematically improving data fidelity.

Materials & methods

Participants

Initial sample

All participants included in this study were enrolled in the PNC (Calkins et al., 2014, 2015; Gur et al., 2014; Satterthwaite et al., 2014). The PNC is a large community-based epidemiological sample of 9498 youths, aged 8–21, who underwent clinical and cognitive evaluations. A subset of 1000 subjects received multimodal neuroimaging as part of the initial PNC project. An additional 601 individuals underwent the identical neuroimaging protocol as part of extension of the PNC. Data from 244 individuals was considered unusable (Fig. 1). Accordingly, 1357 individuals comprise the initial sample that received the same neuroimaging protocol (Table 1). These data were acquired between 2009 and 2012. A description of the PNC is available in: http://www.med.upenn.edu/bbl/projects/pnc/PhiladelphiaNeurodevelopmentalCohort.shtml and the data is available from the National Institutes of Health — dbGaP (http://www.ncbi.nlm.nih.gov/gap).

Validation sample

Four hundred and four individuals (Table 2) returned approximately two years later and underwent the same neuroimaging procedures. Of note, these individuals were selected to return based upon successful completion and high data fidelity of a structural scan during the initial study. DTI quality during the initial study was not a factor in enrollment for follow-up. Thirty individuals did not complete a follow-up DTI scan. Thus, the final sample was 374. These data were collected between 2012 and 2013.

All enrolled subjects provided informed consent at each visit, or for minors informed assent in addition to parental or guardian consent. The Institutional Review Boards of the University of Pennsylvania and Children's Hospital of Philadelphia approved all procedures.

Please cite this article as: Roalf, D.R., et al., The impact of quality assurance assessment on diffusion tensor imaging outcomes in a large-scale population-based cohort, NeuroImage (2015), http://dx.doi.org/10.1016/j.neuroimage.2015.10.068