DTI measures identify mild and moderate TBI cases among patients with complex health problems: A receiver operating characteristic analysis of U.S. veterans

Keith L. Main\textsuperscript{a,b,c,⁎}, Salil Soman\textsuperscript{a,d,e}, Franco Pestilli\textsuperscript{f}, Ansgar Fursa\textsuperscript{a,d,g}, Art Noda\textsuperscript{d}, Beatriz Hernandez\textsuperscript{d}, Jennifer Kong\textsuperscript{a}, Jauhtai Cheng\textsuperscript{a}, Jennifer K. Fairchild\textsuperscript{d}, Joy Taylor\textsuperscript{d}, Jerome Yesavage\textsuperscript{a,d}, J. Wesson Ashford\textsuperscript{b,d}, Helena Kraemer\textsuperscript{d}, Maheen M. Adamson\textsuperscript{d,h,i}

\textsuperscript{a} War Related Illness and Injury Study Center, Veterans Affairs, Palo Alto Health Care System (VAPAHCS), Palo Alto, CA, United States
\textsuperscript{b} Defense and Veterans Brain Injury Center (DVBIG), Silver Spring, MD, United States
\textsuperscript{c} General Dynamics Health Solutions (GDHS), Fairfax, VA, United States
\textsuperscript{d} Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, CA, United States
\textsuperscript{e} Department of Radiology, Stanford University, Palo Alto, CA, United States
\textsuperscript{f} Department of Neurology and Neurological Sciences, Stanford University School of Medicine, Stanford, CA, United States
\textsuperscript{g} Department of Psychology and Brain Sciences, Indiana University, Bloomington, IN, United States
\textsuperscript{h} Department of Neurosurgery, Stanford School of Medicine, Stanford, CA, United States
\textsuperscript{i} Defense and Veterans Brain Injury Center (DVBIG), Veterans Affairs, Palo Alto Health Care System (VAPAHCS), Palo Alto, CA, United States

**ARTICLE INFO**

**Keywords:**
- Traumatic brain injury
- Concussion
- Imaging
- Axon degeneration
- Neurodegeneration

**ABSTRACT**

Standard MRI methods are often inadequate for identifying mild traumatic brain injury (TBI). Advances in diffusion tensor imaging now provide potential biomarkers of TBI among white matter fascicles (tracts). However, it is still unclear which tracts are most pertinent to TBI diagnosis. This study ranked fiber tracts on their ability to discriminate patients with and without TBI. We acquired diffusion tensor imaging data from military veterans admitted to a polytrauma clinic (Overall \( n = 109 \); Age: \( M = 47.2, SD = 11.3; \) Male: 88%; TBI: 67%). TBI diagnosis was based on self-report and neurological examination. Fiber tractography analysis produced 20 fiber tracts per patient. Each tract yielded four clinically relevant measures (fractional anisotropy, mean diffusivity, radial diffusivity, and axial diffusivity). We applied receiver operating characteristic (ROC) analyses to identify the most diagnostic tract for each measure. The analyses produced an optimal cutoff for each tract. We then used kappa coefficients to rate the agreement of each cutoff with the neurologist's diagnosis. The tract with the highest kappa was most diagnostic. As a check on the ROC results, we performed a stepwise logistic regression on each measure using all 20 tracts as predictors. We also bootstrapped the ROC analyses to compute the 95% confidence intervals for sensitivity, specificity, and the highest kappa coefficients. The ROC analyses identified two fiber tracts as most diagnostic of TBI: the left cingulum (LCG) and the left inferior fronto-occipital fasciculus (LIF). Like ROC, logistic regression identified LCG as most predictive for the FA measure but identified the right anterior thalamic tract (RAT) for the MD, RD, and AD measures. These findings are potentially relevant to the development of TBI biomarkers. Our methods also demonstrate how ROC analysis may be used to identify clinically relevant variables in the TBI population.

**Abbreviations:** TBI, traumatic brain injury; DTI, diffusion tensor imaging; ROC, receiver operating characteristic; DAI, diffuse axonal injury; PTSD, post-traumatic stress disorder; FA, fractional anisotropy; MD, mean diffusivity; RD, radial diffusivity; AD, axial diffusivity; LAT, left anterior thalamic tract; RAT, right anterior thalamic tract; LCG, left cingulum; RCG, right cingulum; LCH, left cingulum – hippocampus; RCH, right cingulum – Hippocampus; LCS, left cortico-spinal tract; RCS, right cortico-spinal tract; LIF, left inferior fronto-occipital fasciculus; RIF, right inferior fronto-occipital fasciculus; LIL, left inferior longitudinal fasciculus; RIL, right inferior longitudinal fasciculus; LSL, left superior longitudinal fasciculus; RSL, right superior longitudinal fasciculus; LST, left superior temporal fasciculus – temporal; RST, right superior temporal fasciculus – temporal; LUN, left uncinate; RUN, right uncinate; GN, genu; SP, splenium; CC, corpus callosum

⁎ Corresponding author at: Defense and Veterans Brain Injury Center, 1335 East-West Highway, Suite 4-100, Silver Spring, MD 20910, United States.

E-mail address: keith.l.main3.ctr@mail.mil (K.L. Main).

http://dx.doi.org/10.1016/j.nicl.2017.06.031

Received 15 August 2016; Received in revised form 10 June 2017; Accepted 23 June 2017

Available online 24 June 2017

2213-1582/ © 2017 Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/BY-NC-ND/4.0/).
1. Introduction

In the U.S. military, the rigors of training and dangers of combat yield thousands of traumatic brain injuries (TBI) every year (Defense and Veterans Brain Injury Center, 2012). The Centers for Disease Control and Prevention (2013) reported that from 2000 to 2011, a total of 235,046 service members sustained a TBI, approximately 4% of the armed forces. Traumatic brain injury was also a major source of morbidity during the campaigns in Iraq and Afghanistan (Okie, 2005; Warden, 2006). Many of these veterans have returned to civilian life but continue to endure the symptoms and limitations of their injuries. The military now regards the physical and psychological consequences of TBI as a major medical challenge (Tanielian et al., 2008).

A compelling need exists to develop objective and reliable methods of diagnosing TBI and predicting its outcomes. Over the past decade, diffusion tensor imaging (DTI) has emerged as a promising new technology with both research and clinical implications (Belanger et al., 2007; Dong et al., 2004; Hunter et al., 2012). Measurements derived from DTI sequences assess the integrity of brain tissue by evaluating the diffusion of water molecules (Mulkerjee et al., 2008). Accordingly, DTI may provide a reliable method of detecting microstructural abnormalities in the brain and thereby aid in the development of TBI biomarkers (Bigler, 2013; Bigler and Bazarian, 2010).

In this report, we present a method for exploring relationships between DTI measures and TBI diagnosis. We leveraged a specialized receiver operating characteristic (ROC) analysis to identify white matter tracts and DTI thresholds potentially diagnostic of TBI in U.S. veterans. Our goal was to apply this method in a polytrauma context and assess its findings in relation to the current literature.

1.1. TBI symptoms and diagnosis

Traumatic brain injury is a disruption in normal brain functioning caused by an external force (Menon et al., 2010). The Glasgow Coma Scale (GCS) categorizes TBI cases as mild, moderate, or severe based on the patient’s level of consciousness shortly after injury (Teasdale and Jennett, 1974). The Mild Traumatic Brain Injury Committee of the Head Injury Interdisciplinary Special Interest Group of the American Congress of Rehabilitation Medicine (1993) defines mild TBI (mTBI) as a head injury with an initial GCS score between 13 and 15 (30 min after injury) and at least one of the following criteria: an alteration of consciousness (AOC), such as feeling dazed or confused; a loss of consciousness (LOC), 30 min or less; and posttraumatic amnesia no longer than 24 h after injury.

Diagnosis of TBI has traditionally relied on self-report and neurological testing (Ruff et al., 2009). Initially, patients may experience headache, amnesia, nausea, slurred speech, dizziness, incoordination, and emotional malaise (Kelly and Rosenberg, 1997; Levin et al., 1987). Most will recover in a matter of days, but some experience symptoms weeks, even months later (Alexander, 1995). These lingering sequelae include cognitive, sensory, somatic, and affective components (Cicerone and Kalmar, 1995; Halbauer et al., 2009). Collectively they are referred to as post-concussive syndrome (PCS) (Bigler, 2008; Broshek et al., 2015; Mittenberg and Strauman, 2000).

Scientists continue to debate the nature of PCS (Pricatanato and Gale, 2011). It is unclear, for example, whether the etiology is organic or psychogenic (King, 2003). Most patients will convalesce in one to three months (Levin et al., 1987). However, a proportion (10–30%) will experience symptoms longer (Alexander, 1995; Barlow et al., 2010; Eisenberg et al., 2013; Levin et al., 2013; Ponsford et al., 2011; Vanderploug et al., 2007). This chronic form, known as persistent post-concussional syndrome (PPCS), may trouble some patients indefinitely (McMillan et al., 2012). The burden imposed by PPCS can be substantial, straining a person’s work, relationships, and daily life. Currently, there are no biological or behavioral indicators that signal whether a TBI patient will develop chronic symptoms.

1.2. TBI pathophysiology

Research over the last 20 years has detailed the pathophysiology of TBI (Blennow et al., 2012; Johnson et al., 2013). Moderate and severe TBIs can result in visible intracranial abnormalities: hemorrhages, hematomas, the laceration of cerebrovascular tissues. However, the vast majority of TBIs is closed-head injuries and are mild in nature (Bazarian et al., 2005). Here the damage is microscopic or metabolic. It involves microstructural damage to white matter axons, ionic imbalances, or alterations in neural metabolism (Barkhoudarian et al., 2011).

A common result of TBI is stretching or distortion of white matter axons (Smith and Meaney, 2000). Animal and in vitro models show traumatized axons trigger a biochemical cascade that alters neuronal homeostasis (Giza and Hovda, 2014). Affected neurons experience an abnormal influx of calcium ions into the axon (Wolf et al., 2001) causing mitochondrial swelling (Cheng et al., 2012), neurofilament compaction (Siedler et al., 2014), and compromise of microtubules (Tang-Schomer et al., 2010). This pathology, in turn, impairs axoplasmic transport. Organic materials begin to accumulate along the axon creating periodic swellings (Christman et al., 1994; Pavlishock and Becker, 1985; Smith et al., 1999). These varicosities signal a breakdown in cellular structure (Pettus and Pavlishock, 1996; Pavlishock and Pettus, 1996; Saatman et al., 2003). Axons may eventually disconnect from the cell body and deteriorate, a process known as Wallerian degeneration (Wang et al., 2012).

When widely distributed throughout the brain, this pathology is called diffuse axonal injury (DAI) (Johnson et al., 2013; Smith et al., 2003). Researchers have observed evidence of DAI, post-mortem, for over half a century (Adams et al., 1982; Oppenheimer, 1968; Peerless and Rewcastle, 1967; Strich, 1956). A more complex issue is the condition’s effect on TBI survivors. One hypothesis is DAI impairs communication in large-scale neural networks (Irwinia et al., 2012; Sharp et al., 2014). This disconnection may underlie the cognitive and affective symptoms patients experience post-injury. However, with limited evidence, this is still conjecture. Currently, the only objective biomarkers of DAI are histopathological, but recent advances in neuroimaging hold promise to assess DAI in vivo.

1.3. Diffusion tensor imaging (DTI) analysis

In many cases, the physical damage from mTBI is not detectable using standard, clinical neuroimaging (Bigler and Snyder, 1995; Hughes et al., 2004; Jeret et al., 1993; Lee et al., 2008; Mittil et al., 1994; for a review see Provenzale, 2010). However, DTI has shown potential for identifying microstructural damage in mTBI patients (Bigler, 2013; Bigler and Bazarian, 2010). DTI capitalizes on the nature of water’s diffusion to yield measures of tissue integrity (Basser et al., 1994a, 1994b). White matter bundles are composed of axons, densely packed and collinear in organization. When net diffusion occurs along the fibers’ axis of orientation, it is said to be anisotropic. However, damage to axonal membranes may increase diffusion in all directions, making it more isotropic. By evaluating these properties, DTI provides information on the health and architecture of brain white matter tracts.

All DTI measures are derived from the tensor, a mathematical model that describes the direction and magnitude of diffusion in MRI voxels (Le Bihan et al., 2001; Pierpaoli et al., 1996). Axial diffusivity (AD) is the eigenvalue (magnitude) of the tensor’s principal axis (λ1), which is parallel to fiber orientation. Thus, AD describes the rate of diffusion along a fiber tract’s main pathway. Radial diffusivity (RD), in contrast, describes diffusion in the transverse or perpendicular directions: (λ2 + λ3)/2. Combining these eigenvalues yields other summary measures. Mean diffusivity (MD) describes the total amount of diffusion in a voxel by averaging the parallel (AD) and transverse (RD) eigenvalues: (λ1 + λ2 + λ3)/3. Fractional anisotropy (FA) is the normalized variance of these eigenvalues: \( \sqrt{\frac{3}{2}} \cdot \sqrt{(\lambda_1 - D)^2 + (\lambda_2 - D)^2 + (\lambda_3 - D)^2} / (\lambda_1^2 + \lambda_2^2 + \lambda_3^2) \). It isolates the
summarizes demographics and clinical in-

examine white matter pathology in TBI patients (for reviews see controls (Thomas et al., 2014). In many cases fibers cross, branch, and touch in ways that produce artefactual and/or incomplete reconstructions (Catani, 2007). Still, current methods have clinical utility in evaluating large-scale fiber bundles (Shenton et al., 2003), and tracking of neurodegenerative diseases (Bozzali et al., 2005; Matsuo et al., 2008; Taoka et al., 2006).

Over the past decade, a growing scientific literature has used DTI to examine white matter pathology in TBI patients (for reviews see Eierud et al., 2014; Gardner et al., 2012; Hulkower et al., 2013; Kou et al., 2010; Shenton et al., 2012). Among chronic patients, research consistently shows a reduction in FA and/or increase in MD relative to controls (Cubon et al., 2011; Davenport et al., 2012; Geary et al., 2010; Grossman et al., 2012; Inglese et al., 2005; Kraus et al., 2007; Lipton et al., 2008; Little et al., 2010; Lo et al., 2009; Niogi et al., 2008a, 2008b; Rutgers et al., 2008a, 2008b; Salmond et al., 2006). Research also documents a list of commonly affected tracts including the corpus callosum, cingulum bundle, cortico-spinal tract, superior and inferior longitudinal fasciculus, internal capsule, and the corona radiata (Hulkower et al., 2013; Shenton et al., 2012).

As DTI analysis becomes more refined, it may contribute to the development of TBI biomarkers. However, foundational information is currently lacking. What DTI measures best identify patients with TBI? What brain regions yield pertinent clinical information? Can certain fiber tracts or groups of tracts be linked to specific symptoms? An established technique in medical diagnostics may provide preliminary answers.

1.4. Receiver operating characteristic (ROC) analysis

The receiver operating characteristic (ROC) is a signal detection technique that evaluates the accuracy of a binary classifier on an outcome of interest (Macmillan and Creelman, 2005). Initially developed in the context of radar detection, ROC has been applied in fields as diverse as psychophysics, weather forecasting, and aptitude testing (Swets, 1988). The ROC methodology has also become central to research on medical decision making (Kraemer, 1992; Metz, 1986; Zweig and Campbell, 1993). For example, an exploratory ROC technique has recently proven useful in identifying the characteristics of clinical subgroups at risk for certain outcomes (Fairchild et al., 2013; Hoblyn et al., 2006; Kiernan et al., 2001; Kinoshita et al., 2012; Noda et al., 2006; O’Hara et al., 2002a; Periyakoll et al., 2005; Thanassi et al., 2012; Tinklenberg et al., 2015; Yesavage et al., 2003). The method has revealed predictors of rapid cognitive decline in Alzheimer’s disease patients (O’Hara et al., 2002b), the onset of depression in the seriously ill (Periyakoll et al., 2012), and the decline of flight simulator performance in aging pilots (Yesavage et al., 2011).

The advent of extensive patient databases and personalized interventions will likely make ROC just as applicable in the future (Alemayehu and Zou, 2012). New approaches in precision medicine profile patients based on their risk of a disease, potential for recovery, and interaction with treatments (Mirmezami et al., 2012). The idea is to acknowledge heterogeneity in patient populations and tailor treatments accordingly. The bedrock of this approach is a data-driven taxonomy. Exploratory ROC may prove invaluable in this regard. It has the potential to make connections between test results, patient profiles, and likely outcomes. Thus, in a complex domain like brain injury, ROC may distil and identify crucial information.

1.5. Current study

We submit there is merit in using ROC to explore the diagnostic accuracy of DTI measures. While cross-sectional research has identified group differences between TBI cases and controls, these findings have little to no clinical significance. They do not inform the diagnosis and/or classification of individual patients. There is a need then to apply methods in clinical diagnostics to DTI data. Doing so may identify measures, thresholds, and brain regions clinically relevant to the diagnosis and care of TBI patients.

In the current study, we applied ROC analysis to DTI measures. We acquired DTI scans from a sample of veterans evaluated at War Related Illness and Injury Study Center (WRIISC). The patients were diagnosed with or without TBI based on VA/DoD guidelines. DTI pre-processing and tractography analysis generated a set of twenty major fiber tracts for each patient. Standard measures (FA, MD, RD, and AD) were extracted from these tracts and used as predictors in a set of exploratory ROC analyses.

2. Material and methods

2.1. Patients

Patients were assessed at the War Related Illness and Injury Study Center (WRIISC-CA), a VA second-level clinic dedicated to performing comprehensive evaluations on veterans with post-deployment health concerns (Lange et al., 2013). Funded by the Office of Public Health, WRIISC sites (East Orange, NJ, Washington, DC, and Palo Alto, CA) accept veteran patients with complex medical and behavioral conditions including TBI and PTSD. Stanford University’s IRB committee and the VA Palo Alto Health Care System’s (VAPAHCs) research administration approved the study. Patient consent was obtained according to the Declaration of Helsinki.

We recruited 109 veteran patients (Age: $M = 47.2, SD = 11.3$; Male: 88%) from the WRIISC clinic at VAPAHCs. Ninety-seven of these patients were previously deployed (Operation Iraqi Freedom = 30, Operation Enduring Freedom = 13, Somalia = 4, Bosnia = 2, Gulf War = 56, Grenada = 1, Vietnam = 15, Korea = 2, and Other = 2). The sum of 125 is due to multiple deployments. All patients received a comprehensive physical, neurological, neuropsychological, and psychiatric evaluation. Exclusion criteria included schizophrenia, bipolar and psychotic disorders, somatoform disorders, and a history of neurological disorder (other than PTSD) not caused or precipitated by traumatic injury. Seventy-three of the patients were diagnosed with TBI. Among the TBI patients, 15 (21%) individuals were diagnosed with only TBI; 57 (78%) had both TBI and PTSD. One TBI patient lacked a PTSD diagnosis. Thirty-six of the patients were not diagnosed with TBI. Among the non-TBI patients, 21 (58%) were diagnosed with PTSD, 15 (42%) without. Table 1 summarizes demographics and clinical information for the patients.

Sixty-three (86%) of the TBI cases were classified as mild, 10 (14%) as moderate. There were no severe cases. Most TBI patients reported some loss of consciousness (LOC). Forty-four (60%) had an LOC with a duration under 24 h. Eighteen (25%) reported no LOC. Data was not available on 11 (15%) patients. Forty-four (60%) of the TBI patients reported injuries due to impact, 9 (12%) from blast exposure, and 20 (27%) from a combination of impact and blast. An exact time since injury was not available given our reliance on self-report data. We do know patients deployed before OEF/OIF sustained their injuries several years ago. In addition, the process of primary care, referral, and scheduling means even OEF/OIF patients were months, if not years, post-injury when admitted to WRIISC.

Finally, we compared age, years of education, the number of diagnoses, and the number of medications between the TBI and non-TBI groups. The TBI ($M = 18.6, SD = 9.7$) patients had significantly more diagnoses than the non-TBI ($M = 13.1, SD = 8.1$) patients, t(98)
Table 1
Patient demographics and clinical characteristics.

<table>
<thead>
<tr>
<th></th>
<th>With TBI (n = 73)</th>
<th>Without TBI (n = 36)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis: TBI + PTSD</td>
<td>57 (78%)</td>
<td>0</td>
</tr>
<tr>
<td>TBI only</td>
<td>15 (21%)</td>
<td>0</td>
</tr>
<tr>
<td>PTSD only</td>
<td>0</td>
<td>21 (58%)</td>
</tr>
<tr>
<td>Neither TBI nor PTSD</td>
<td>0</td>
<td>15 (42%)</td>
</tr>
<tr>
<td>Symptomatic Cognitive</td>
<td>43 (61%)</td>
<td>13 (42%)</td>
</tr>
<tr>
<td>Pain</td>
<td>68 (97%)</td>
<td>29 (94%)</td>
</tr>
<tr>
<td>Fatigue</td>
<td>30 (43%)</td>
<td>13 (42%)</td>
</tr>
<tr>
<td>Sleep</td>
<td>60 (86%)</td>
<td>26 (84%)</td>
</tr>
<tr>
<td>Pulmonary</td>
<td>19 (27%)</td>
<td>8 (26%)</td>
</tr>
<tr>
<td>Dermatological</td>
<td>25 (36%)</td>
<td>9 (29%)</td>
</tr>
<tr>
<td>Gastrointestinal</td>
<td>44 (63%)</td>
<td>18 (58%)</td>
</tr>
<tr>
<td>Other</td>
<td>70 (100%)</td>
<td>29 (94%)</td>
</tr>
<tr>
<td>Total, M (SD)</td>
<td>18.6 (9.7)</td>
<td>13.1 (8.1)</td>
</tr>
<tr>
<td>Medications, M (SD)</td>
<td>9.8 (7.4)</td>
<td>8.3 (6.9)</td>
</tr>
<tr>
<td>Age in years, M (SD)</td>
<td>47.7 (12.0)</td>
<td>46.3 (9.7)</td>
</tr>
<tr>
<td>Years of education, M (SD)</td>
<td>14.6 (2.5)</td>
<td>13.7 (2.6)</td>
</tr>
<tr>
<td>Men, n (%)</td>
<td>63 (86%)</td>
<td>33 (92%)</td>
</tr>
</tbody>
</table>

Note: Diagnosis: One participant with TBI lacked a PTSD diagnosis. Therefore, TBI only and TBI + PTSD percentages from this column do not add to 100%. Symptoms: Patients had multiple symptoms. Accordingly, column percentages add to over 100%. Symptom data was not available for 3 patients with TBI, 5 patients without TBI.

= −2.74, p = 0.007. Age, years of education, and the number of medications were not significantly different between the groups.

2.2. Patient diagnosis

Board-certified neurologists and clinical psychologists determined diagnoses for the TBI and PTSD patients, respectively. Physicians based the TBI diagnosis on standards from the American Congress of Rehabilitation Medicine Head Injury Interdisciplinary Special Interest Group. WRIISC neurologists conducted a structured interview for each patient. The questions evaluated the incidence and circumstances of TBI across the lifespan, not just during military service. Patients were questioned on persistent symptoms, including headaches, paralysis, dizziness, vertigo, fatigue, mobility, sleep disturbances, as well as cognitive and emotional problems. Physical examinations assessed muscle tone, reflexes, gait, spasticity, sensory function, and cerebellar signs. The clinical team met and evaluated each case, discussing the interview, military records, physical examination, and neuroimaging (CT, PET, and structural MRI). For most patients, CT and structural MRI images were negative. A prior review of WRIISC records found that approximately 10% of scans showed an indication of brain injury (e.g., hemorrhage, edema, ischemia, microbleeds). Accordingly, self-report and clinical observations were the main determinants of TBI diagnosis and severity.

Psychologists used the Clinician-Administered PTSD Scale (CAPS) (Blake et al., 1995) with guidelines set forth by the International Society for Traumatic Stress Studies to diagnose PTSD. Through a semi-structured interview the CAPS assesses behavioral features of PTSD as defined by the DSM-IV (American Psychiatric Association, 2000). The interviewer identified a traumatic event or events to serve as a basis for inquiry then assessed PTSD symptoms using 30 standard questions and probes. The questions targeted 5 PTSD criteria: re-experiencing, avoidance of stimuli, persistent arousal, duration of symptoms, and impact on social and professional life. Questions specifically evaluated current symptoms. The interviewer rated all responses to initial and follow-up questions then made diagnostic and severity determinations based on established scoring rules and observed patient behavior. The CAPS took between 30 and 60 min to administer.

2.3. MRI acquisition

Neuroimaging was conducted at VAPAHCS. We acquired all MRI data on a 3 T Discovery MR750 (GE Medical Systems, Milwaukee, WI) with an eight-channel, GE head coil. Each neuroimaging session lasted approximately 40 min and included a DTI sequence and several structural scans (T1 and T2-weighted, SWI, FLAIR). The current analyses exclude the structural scans. We acquired the DTI data using a single-shot, spin-echo, echo-planar sequence with 50 axial slices (TE = 80 ms; TR = 6600 ms; FOV = 240 mm; slice thickness = 2.5 mm, no gap; acquisition matrix = 96 × 96; number of excitations = 1). Diffusion weighting was applied along 30 non-collinear directions (b = 1000 s/mm²). The sequence had 5 non-diffusion (b = 0) volumes. We collected two DTI acquisitions run at NEX = 1 and combined them in post-processing to improve the signal-to-noise ratio. The reconstructed voxel dimensions were 2.5 × 2.5 × 2.5 mm.

2.4. MRI analysis

2.4.1. Preprocessing

We analyzed DTI data using mrDiffusion, open-source software developed by the Vista Laboratory, Stanford University (http://white.stanford.edu/software). Images from each patient’s two DTI sequences were concatenated, and the resulting file entered a preprocessing pipeline consisting of bias correction (motion, eddy current, and EPI distortion), non-linear registration to the Montreal Neurological Institute (MNI) template, resampling, and fitting of the tensor model. We subsequently examined head motion in the TBI and non-TBI patients. The average motion in both groups was less than one voxel (TBI: x = 0.91, y = 0.56, z = 0.66; non-TBI: x = 0.79, y = 0.52, z = 0.52) or < 2.5 mm.

Fiber tracking was carried out using a customized version of MRtrix software (Tournier et al., 2012). MRtrix is an open-source suite of tools for analysis of diffusion MRI data. Its tractography algorithms use constrained spherical deconvolution and probabilistic streamline tracking to produce fibers robust against crossing fiber effects. We incorporated MRtrix in our processing stream using procedures described in Pestilli et al. (2014). Fiber tracking with MRtrix produced a whole-brain tractography for each patient.

2.4.2. Tractography

The whole-brain tractography is a representation of the brain’s white matter connectome. In these models, the direction of diffusion is tracked voxel-to-voxel, producing virtual “fibers” that connect brain regions. These fibers have a coherent organization and may be grouped together into large-scale fascicles or tracts. A way of doing this is to define two or more regions of interest (ROIs) in specific parts of the brain then identify all fibers that pass through these areas (Wakana et al., 2004). Segregating these fibers as a collective yields tracts with good agreement to gross level neuroanatomy (Wakana et al., 2007). We employed the open-source software Automated Fiber Quantification (AFQ) to segment patient tractographies (http://github.com/jyeatman/AFQ). This automated pipeline uses a two-ROI approach and produces tracts highly correlated with manual segmentation (Yeatman et al., 2012).

For automated segmentation, AFQ used pre-defined ROIs from a DTI data set registered to the MNI template. The AFQ developers drew these ROIs based on prescriptions from Wakana et al. (2007). We mapped the whole-brain tractography of each patient onto this atlas using a non-linear version of the transformation described in Zhang et al. (2008). Unique ROI pairs served as waypoints for each tract of interest. AFQ isolated individual tracts by subtracting all fibers that failed to traverse both ROIs.

The pipeline further refined tracts through a two-step cleaning process. (1) AFQ made voxel-wise comparisons between fiber trajectories and tract probability maps. Hua et al. (2008) manually
segmented and coregistered tracts from a group of healthy adults. For every tract, they calculated a probability map depicting its likely trajectory. Applying these maps to the current data, AFQ culled every tract, they calculated a probability map depicting its likely trajectory. AFQ removed fiber outliers in relation to each tract’s core. Tract cores were generated by dividing the length of each tract into 90 equidistant sections then calculating the mean of the fibers’ x, y, and z coordinates for every section. The Mahalanobis distance of a fiber from the section’s core was interpreted as a z-score on a 3-dimensional, Gaussian distribution. AFQ removed fibers > 3 standard deviations from the core mean. This process was iterated across sections until no outliers remained.

In a final check, visual inspection and manual editing by trained staff removed any remaining aberrant fibers. The segmentation and cleaning process rendered 20 fiber tracts for each patient (Fig. 1): the left and right anterior thalamic radiations (LAT and RAT), left and right cingulum (LCG and RCG), left and right cingulum-hippocampus (LCH and RCH), left and right cortico-spinal tract (LCS and RCS), left and right inferior fronto-occipital fasciculus (LIFO and RIFO), left and right inferior longitudinal fasciculus (LIL and RIL), left and right superior longitudinal fasciculus (LSL and RSL), left and right superior longitudinal fasciculus-temporal (LST and RST), and left and right uncinate fasciculus (LUN and RUN), the genu (GN), and splenium (SP). More information on AFQ’s procedures for tract segmentation and editing can be found in Yeatman et al. (2012).

2.4.3. DTI quantification

In-house, custom MATLAB code quantified fiber diffusion properties for every tract. These measures include fractional anisotropy (FA), mean diffusivity (MD), radial diffusivity (RD), and axial diffusivity (AD). Each tract was previously divided into 90 sections. The program calculated the weighted average of each measure for each section. An individual fiber’s contribution to this value was determined by calculating its Mahalanobis distance from the core tract (Yeatman et al., 2012). Thus fibers farther away from the core were weighted less in the averages. This procedure reduced the influence of partial volume effects, where adjacent tissues (gray matter, cerebral spinal fluid, other white matter) bias tract diffusion measures. Finally, a representative value for each tract was determined by calculating the arithmetic mean of all its sections.

2.5. Statistical analysis

We conducted a three-part statistical analysis on the data. (1) Exploratory ROCs identified tracts predictive of the TBI diagnosis for each measure (FA, MD, RD, and AD). (2) We implemented a resampling procedure on these ROCs, generating confidence intervals around their outputs. (3) We then applied a stepwise logistic regression to each measure with the purpose of verifying the ROC-identified tracts.

2.5.1. Receiver operating characteristic analysis

We used publicly available software to conduct the ROC analyses (ROC version 4.22; http://www.stanford.edu/~yesavage/ROC.html). The software was specially designed to evaluate clinical databases for predictive variables. This application of ROC is non-parametric and non-hypothesis testing. It accepts any type of data (dichotomous, ordinal, and continuous) and has no limit on the number of variables assessed. It is well-suited to hypothesis generation in scenarios where the diagnosis might be influenced by many variables and data that are characterized by collinearity and interaction effects. We ran a separate ROC for each DTI measure (FA, MD, RD, and AD). Patient age and the 20 fiber tracts were the independent variables or predictors. The dependent or outcome variable for each analysis was TBI diagnosis.

The ROC searches all predictor variables and selects the optimal classifier for a data set. This predictor has the best agreement with the outcome variable. The optimal classifier is determined by calculation of Cohen’s kappa (Cohen, 1960), a measure of inter-rater agreement for categorical items. The ROC program ranks all variables according to the size of their kappa coefficients and selects the variable with the largest as the optimal classifier. The selection must also demonstrate a significant chi-square statistic (p < 0.05), indicating a relationship between the predictor and outcome variables.

The ROC software addresses two types of data in its calculations. When all variables in the data set are binary (e.g., male/female), it simply calculates the kappa coefficients and selects the largest among them. When one or more variables are ordinal or continuous (e.g., age; severity), identifying the optimal classifier involves two steps. (1) The program first determines data cutpoints. For each variable it reorders the values with respect to magnitude then divides the data at a point that produces the best agreement with the outcome variable. (2) The program then compares all these dichotomized variables and determines which one has the largest kappa coefficient.

Once an optimal classifier is selected, the ROC program divides the data according to its prescribed grouping. The analysis then repeats and kappa coefficients are calculated and ranked for the two new groups. If another classifier is identified in either group, they will divide again. The algorithm will continue this way until a sub-group is too small for further analysis (n < 20), the chi-square is not significant (p > 0.05), or there are no further variables to select. The ROC essentially performs a type of recursive partitioning, dividing the sample into smaller and
smaller subsets as long as it can identify a classifier and stopping rules do not occur. The end result is a decision tree that arranges classifiers in a hierarchy of diagnostic importance. The program also produces a number of values critical to inter-rater assessment. For each classifier, these include its kappa coefficient, sensitivity, specificity, and cutpoint (if applicable).

2.5.2. Resampling analysis

To test the accuracy of the initial ROCs, we addressed whether the tracts are frequently identified as optimal classifiers when the analysis is iterated. To do so, we implemented a three-step resampling procedure using custom code in SAS (http://www.sas.com). This analysis also generated confidence intervals for the tract cutpoints, sensitivity, specificity, and kappa coefficients.

(1) For each DTI measure, we resampled with replacement 1000 times from the original data and performed a ROC analysis on each sample. We tallied the number of times each tract was identified as the optimal classifier. Tracts identified over 100 times were kept for further analysis.

(2) To assess collinearity, we performed Spearman correlations and a principal component (PCA) analysis on the tracts from Step 1. We first determined whether any of the tracts are highly correlated with the initial optimal classifiers. Tracts with Spearman correlation coefficients > 0.8 were discarded. The rest of the tracts were evaluated by PCA. We kept for further analysis only those tracts that loaded onto the first principal component.

(3) Using the remaining variables from Step 2, we re-ran the resampling analysis on each measure. As before, the procedure yielded 1000 unique samples with a ROC on each. From these results, we determined our final set of classifiers (reported in the Results). For each measure, we selected the tract identified as the optimal classifier in the majority (> 500) of samples. Once the optimal classifiers were selected, we calculated 95% confidence intervals (CI) for their cutpoints, sensitivity, specificity, and kappa coefficients (Table 2). To do this, we aggregated the results for each test statistic into sampling distributions and selected the 97.5th and 2.5th centiles as upper and lower bounds, respectively. Thus, for each sampling distribution, the range of values representing the confidence interval has a 95% chance of containing the population parameter.

2.5.3. Stepwise logistic regression

As a confirmation of the ROC findings, we applied stepwise logistic regression to each of the DTI measures. Stepwise logistic regression is appropriate here given the large number of predictor variables and exploratory nature of analyses. Each regression involved the entire data set (n = 109). The 20 tracts served as predictors for each of the four analyses. TBI diagnosis was the dependent variable.

3. Results

3.1. Initial exploratory ROC

Our initial ROC analyses identified two primary tracts of interest. The left cingulum (LCG) was the optimal classifier for the FA measure. The left inferior fronto-occipital fasciculus (LIF) was optimal classifier for the MD, RD, and AD measures. In addition, the MD ROC yielded a secondary classifier in the left superior longitudinal fasciculus – temporal (LST). For the other ROCs, stopping rules halted the analysis after the first division.

3.2. Resampling and identification of optimal classifiers

In Step 1 of the resampling analysis, we selected tracts identified as optimal classifiers over 100 times. These included LCG and RIL for the FA measure; LIF and RCG for the MD measure; LIF, LAT, RAT, and RIL for the RD measure; and LIF, LCG, and RCG for the AD measure. In Step 2, we found that none of the newly identified tracts (i.e., RIL, RCG, LAT, RAT, RIL, and RCG) were highly correlated with the optimal classifiers (LCG, LIF). Principal components analysis for each measure yielded only one factor with an eigenvalue above Kaiser’s criterion (λ > 1). All tracts clustered onto the first principal component. Finally, in Step 3, the resampling analysis demonstrated that the majority tracts (i.e., identified as optimal classifiers > 500 times) were the same as those identified in the initial ROCs: LCG for the FA measure; LIF for the MD, RD, and AD measures.

3.3. Fractional anisotropy (FA)

The ROC analysis of the FA data demonstrated that the left cingulum bundle (LCG) has the highest significant kappa (κ = 0.264, χ² = 7.60, p < 0.01). In relation to TBI diagnosis, LCG’s sensitivity is 74%. Specificity is 52.8% (Table 2, Fig. 2a). The optimal cut point for LCG is a mean FA value of 0.433. Of the 71 patients with FA < 0.433, 54 (76.1%) have TBI. Of the 38 patients with FA greater than or equal to 0.433, 19 (50%) have TBI (Fig. 3a). A logistic regression analysis on the whole sample confirmed that LCG significantly predicts TBI status (χ² = 6.62, p < 0.05 with df = 1; Table 3).

Table 2

Receiver operating characteristic (ROC) results.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Tract</th>
<th>Cutpt.</th>
<th>Sens.</th>
<th>Spec.</th>
<th>κ</th>
<th>χ²</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>FA</td>
<td>LCG</td>
<td>0.433</td>
<td>74.0%</td>
<td>52.8%</td>
<td>0.264</td>
<td>7.60</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.397–0.452)</td>
<td>(44.0–89.5%)</td>
<td>(37.1–89.7%)</td>
<td>(0.212–0.493)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MD*</td>
<td>LIF</td>
<td>0.837</td>
<td>68.5%</td>
<td>61.1%</td>
<td>0.278</td>
<td>8.72</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.820–0.861)</td>
<td>(54.8–88.0%)</td>
<td>(35.3–80.6%)</td>
<td>(0.139–0.478)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RD</td>
<td>LIF</td>
<td>0.613</td>
<td>74.0%</td>
<td>55.6%</td>
<td>0.289</td>
<td>9.15</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.610–0.647)</td>
<td>(52.1–83.8%)</td>
<td>(41.9–78.1%)</td>
<td>(0.140–0.476)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AD</td>
<td>LIF</td>
<td>1.272</td>
<td>72.6%</td>
<td>52.8%</td>
<td>0.249</td>
<td>6.76</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.272–1.314)</td>
<td>(53.8–84.8%)</td>
<td>(44.7–83.9%)</td>
<td>(0.182–0.478)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*ROC secondary cutpoint (where LIF is < 0.837, n = 45)

Note. Cutpt, cutpoint; Sens, sensitivity; Spec, specificity; κ, Cohen’s kappa; FA, fractional anisotropy; MD, mean diffusivity; RD, radial diffusivity; AD, axial diffusivity; LCG, left cingulum tract; LIF, Left Inferior Fronto-Occipital Fasciculus; Values in parentheses represent the lower and upper bounds (i.e., Lower – Upper) of the 95% CI.
3.4. Mean diffusivity (MD)

The ROC of the MD data found that the left inferior fronto-occipital fasciculus (LIF) has the highest significant kappa ($\kappa = 0.278$, $\chi^2 = 8.72$, $p < 0.01$). LIF sensitivity is 68.5%. Specificity is 61.1% (Table 2, Fig. 2b). The optimal cutpoint for LIF is a mean MD value of 0.837. Of the 64 patients with MD greater than or equal to 0.837, 50 (78.1%) have TBI. Of the 45 patients with MD < 0.837, 23 (51.1%) have TBI (Fig. 3b).

In addition, the ROC identified another classifier, the left superior longitudinal fasciculus – temporal (LST), among the 45 cases with MD values < 0.837 ($\kappa = 0.379$, $\chi^2 = 6.51$, $p < 0.05$). LST sensitivity is 65.2%, specificity 72.7% (Table 2, Fig. 2b). The optimal cutpoint for LST is a mean MD value of 0.744. Of the 24 patients with MD greater than or equal to 0.744, 8 (33.3%) have TBI. Of the 21 patients with MD < 0.744, 15 (71.4%) have TBI (Fig. 3b). Logistic regression analysis on the whole sample did not confirm the ROC but identified RAT as the most predictive tract ($\chi^2 = 5.68$, $p < 0.05$ with df = 1; Table 3).

3.5. Radial diffusivity (RD)

The ROC of the RD data demonstrated that the left inferior fronto-occipital fasciculus (LIF) has the highest significant kappa ($\kappa = 0.289$, $\chi^2 = 9.15$, $p < 0.01$). Tract sensitivity is 74%, specificity 55.6% (Table 2, Fig. 2c). The optimal cutpoint for LIF is a mean RD value of 0.613. Of the 70 patients with RD greater than or equal to 0.613, 54 (77.1%) have TBI. Of the 39 patients with RD < 0.613, 19 (48.7%) have TBI (Fig. 3c). Logistic regression analysis on the whole sample did not confirm the ROC but identified RAT as the most predictive tract ($\chi^2 = 3.86$, $p < 0.05$ with df = 1; Table 3).

3.6. Axial diffusivity (AD)

The ROC of the AD data showed that the left inferior fronto-occipital fasciculus (LIF) has the highest significant kappa: ($\kappa = 0.249$, $\chi^2 = 6.76$, $p < 0.01$). Tract sensitivity is 72.6%, specificity 52.8% (Table 2, Fig. 2d). The optimal cutpoint for LIF is a mean AD value of 1.272. Of the 70 patients with AD greater than or equal to 1.272, 53 (75.7%) have TBI. Of the 39 patients with AD < 1.272, 20 (51.3%) have TBI (Fig. 3d). A logistic regression analysis on the whole sample did not confirm the ROC but identified RAT as the most predictive tract: ($\chi^2 = 3.86$, $p < 0.05$ with df = 1; Table 3).

4. Discussion

We used exploratory ROC analysis to assess the agreement between DTI measures and TBI diagnosis. Our procedure involved ROCs...
performed on four DTI measures. They identified the left cingulum bundle (LCG) as the optimal classifier for the FA measure and the left inferior fronto-occipital fasciculus (LIF) for the MD, RD, and AD measures. We obtained these results from both the initial ROCs and the resampling analyses. For each measure, we calculated 95% CIs around the optimal classifier’s sensitivity, specificity, cutpoint, and kappa coefficient. Finally, we employed stepwise logistic regression to independently verify these tracts as predictors. The regression results showed that LCG was the best predictor for the FA measure. However, RAT, as opposed to LIF, was the best predictor for the MD, RD, and AD measures.

4.1. Optimal classifiers

The cingulum runs longitudinally above the corpus callosum, connecting the frontal lobe and the parahippocampal gyrus of the temporal lobe (Wakana et al., 2004). The inferior fronto-occipital fasciculus innervates the occipital cortex above the optic radiations and runs laterally and inferiorly to the anterior tip of the temporal lobe (Wakana et al., 2004). Both tracts are identified in cross-sectional DTI studies comparing TBI patients with the neurologically healthy. A recent review by Hulkower et al. (2013) found that 23 out of 100 DTI papers report a compromised cingulum in TBI patients. 12 report the inferior fronto-occipital fasciculus.

Pathology in the cingulum and inferior fronto-occipital fasciculus is associated with deficits in memory and executive function. Kraus et al. (2007) examined DTI measures and neuropsychological performance in patients with chronic TBI. They found FA from the cingulum and inferior fronto-occipital fasciculus are negatively correlated with memory and executive domain scores. More recently, studies on U.S. veterans with mild TBI report FA from the cingulum is associated with impaired executive function (Sorg et al., 2014) and is negatively correlated with processing speed (Sorg et al., 2015). Among brain damaged patients (i.e., stroke and/or TBI), FA from the inferior fronto-occipital fasciculus is correlated with performance on semantic tasks (Han et al., 2013). In sum, the tracts identified by our ROCs are frequently identified in the DTI literature and are associated with known TBI pathology.

Confirmation of these results with stepwise logistic regression is mixed. For the FA measure, the regression analysis found that LCG is most predictive of TBI diagnosis. However, the same analysis identified the right anterior thalamic tract (RAT) for the MD, RD, and AD measures. These results differ from those of the ROC, which found LIF was the optimal classifier. There could be many reasons for this discrepancy. Our ROC analysis was limited to only one division. With more participants, RAT could emerge as a classifier at later points in the decision tree. Another consideration is that the best predictor from the stepwise regression must satisfy its linear parameters. The ROC analysis is not constrained in this way.

However, our regression results are not unique to the TBI/DTI literature. Other DTI studies have demonstrated abnormal FA, MD, and RD values in the anterior thalamic tracts of TBI patients (Dennis et al., 2015; Kinnunen et al., 2011). In addition, anterior thalamic measures account for variance on neuropsychological assessments of memory, attention, semantic processing, and executive function (Han et al., 2013; Little et al., 2010).

It is possible then the ROC and regression analyses identified two different, but clinically relevant tracts. Inherent differences in the analyses may be the cause. A study by Kiernan et al. (2001) reports one such discrepancy. They used ROC and logistic regression to identify high-risk individuals in a patient population. While the analyses identified similar predictors, they did not classify cases in the same way. The ROC grouped patients with homogenous outcomes and homogeneous

---

**Table 3**

<table>
<thead>
<tr>
<th>Measure</th>
<th>Parameter</th>
<th>df</th>
<th>Parameter estimate</th>
<th>Standard error</th>
<th>Wald $\chi^2$</th>
<th>Pr &gt; $\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>FA</td>
<td>Intercept</td>
<td>1</td>
<td>5.86</td>
<td>2.03</td>
<td>8.3</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>LCG</td>
<td>1</td>
<td>$-12.41$</td>
<td>4.82</td>
<td>6.62</td>
<td>0.010</td>
</tr>
<tr>
<td>MD</td>
<td>Intercept</td>
<td>1</td>
<td>$-9.62$</td>
<td>4.4</td>
<td>7.8</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td>RAT</td>
<td>1</td>
<td>12.29</td>
<td>5.27</td>
<td>4.44</td>
<td>0.019</td>
</tr>
<tr>
<td>RD</td>
<td>Intercept</td>
<td>1</td>
<td>$-7.88$</td>
<td>3.58</td>
<td>5.85</td>
<td>0.027</td>
</tr>
<tr>
<td></td>
<td>RAT</td>
<td>1</td>
<td>13.26</td>
<td>5.56</td>
<td>6.58</td>
<td>0.017</td>
</tr>
<tr>
<td>AD</td>
<td>Intercept</td>
<td>1</td>
<td>$-7.83$</td>
<td>4.33</td>
<td>3.27</td>
<td>0.070</td>
</tr>
<tr>
<td></td>
<td>RAT</td>
<td>1</td>
<td>6.96</td>
<td>3.54</td>
<td>3.86</td>
<td>0.049</td>
</tr>
</tbody>
</table>

Note. FA, fractional anisotropy; MD, mean diffusivity; RD, radial diffusivity; AD, axial diffusivity; LCG, left cingulum tract; RAT, right anterior thalamic tract.
predictors together. Logistic regression, however, yielded groups with homogeneous outcomes, but heterogeneous predictors. Because of ROC's ability to identify homogeneous subgroups, Kiernan et al. (2001) suggest it may be more useful in exploring patient risk factors than regression.

One tract not identified in either analysis was the corpus callosum (CC). This interhemispheric fiber bundle is one of the most commonly damaged tracts in TBI patients (for reviews see Hulker et al., 2013; Shenton et al., 2012). A recent meta-analysis of DTI studies from Aoki et al. (2012) reports that the splenium, the most posterior segment of the CC, is frequently compromised after TBI. For this analysis, we included both the splenium (SP) and the anterior genu (GN) in the ROC. Neither were optimal classifiers. More detailed analyses, parsing the body of the commissure, may yield different results.

However, it is important to consider that the tracts frequently identified in cross-sectional studies may not necessarily make the best classifiers. For any tract, the extent of agreement between DTI measures and a diagnosis may be attributable to the frequency and magnitude of damage as well as normal variance in the DTI scores. Tract size, location, and the presence of crossing fibers may all be factors. For example, because the cingulum bundle connects the frontal and temporal lobes, it may be especially vulnerable to contrecoup injury (Gogio, 1941). Here a jolt or impact to the back of the head causes the brain to move inside the skull, hitting the anterior cingulum (Bigler, 2007). Similarly, the inferior fronto-occipital fasciculus is one of the largest tracts in the brain. Its size and location make it subject to the biomechanics of lateral, blunt-force trauma, arguably the most common cause of TBI (Delaney et al., 2006; Hodgson et al., 1983). The frequency of these injuries may make LCG and LIF better classifiers, even if the corpus callosum has more pronounced pathology.

At the moment, these are informed speculations. However, the ROC results do have an accord with the TBI/DTI literature. We submit the identification of LCG and LIF as successful classifiers has a basis in actual pathology. This information may inform future research as well as efforts to develop TBI biomarkers. Ultimately, though, it is still unclear what parameters make one tract a better classifier than another.

4.2. FA and MD measurements

Fractional anisotropy (FA) is a DTI measure widely applied in investigations of TBI. It quantifies the degree of anisotropy in the diffusion process (Basser and Pierpaoli, 1996). Values between 0 and 1 denote the fraction of diffusion along the tensor's principal axis, as opposed to the transverse directions. Higher FA values indicate constrained diffusion. In the case of the brain, this is often due to collinear cellular structures, as found in white matter fascicles (Mukherjee et al., 2008). Lower FA values indicate relatively unconstrained diffusion, characteristic of gray matter and cerebral spinal fluid.

In contrast, mean diffusivity (MD) quantifies the overall degree of diffusion in tissue (Basser and Pierpaoli, 1996). The measure represents the average displacement of water molecules independent of fiber orientation. Higher MD values are associated with neuropathology because disrupted or distorted fiber bundles allow a greater influx of water across cellular membranes (Gass et al., 2001). Brain tissue with high MD values may have potential inflammation, edema, necrosis, or neoplasia (Alexander et al., 2007).

A growing literature links TBI pathology to abnormally low FA and high MD values. Our results reinforce these findings. The ROC found that lower FA values from the LCG (< 0.433) are associated with the TBI diagnosis, which is consistent with several published reports (Bendlin et al., 2008; Bigler et al., 2010; Bonnelle et al., 2011; Kinnunen et al., 2011; Kraus et al., 2007; Levin et al., 2011; Mac Donald et al., 2011; Marquez de la Plata et al., 2011; Maruta et al., 2010; Niogi et al., 2008a; Pal et al., 2012; Palacios et al., 2011; Rutgers et al, 2008b; Sugiyama et al., 2009; Wilde et al., 2010; Xu et al., 2007; Yurgelun-Todd et al., 2011).

The ROC also demonstrated that higher MD values (> 0.837) from the LIF are associated with TBI diagnosis. Abnormally high MD values are also common in TBI patients (Chu et al., 2010; Henry et al., 2011; Lipton et al., 2008; Lipton et al., 2009; Marquez de la Plata et al., 2011; Perlberg et al., 2009; Porto et al., 2011; Salmond et al., 2006). Other investigations report high MD from the LIF specifically (Bendlin et al., 2008; Kinnunen et al., 2011; Messe et al., 2011; Xu et al., 2007).

While most DTI studies associate low FA values with TBI, a handful of reports say the same for high FA (Bazarian et al., 2007; Henry et al., 2011; Ling et al., 2012; Mayer et al., 2010; Wilde et al., 2008). The literature's discrepancy regarding high and low FA may stem from differences in acute and chronic injury. Findings of high FA in TBI patients may be limited to acute cases where cytotoxic edema has caused an influx of water into axons' intracellular space (Ling et al., 2012; Mayer et al., 2010). In contrast, low FA may characterize a later stage of neuropathology involving demyelination and axonal degeneration.

The current findings seem consistent with this explanation. Our patients were imaged months to years after injury. Accordingly, their chronic condition may preclude finding high FA as a biomarker. We argue our findings only have implications for understanding long-term pathology. Further studies are necessary to examine the diagnostic value of DTI measures at acute and sub-acute stages.

4.3. RD and AD measurements

The ROC analysis demonstrated that the LIF has the highest significant kappa value for both the RD and AD measures. Higher RD (> 0.613) and AD values (> 1.272) from this tract are associated with TBI diagnosis. The relationship of these measures to TBI is less clear. Previous work suggests that RD and AD may be more precise measures of neural pathology because they involve specific eigenvalues (Alexander et al., 2007). The RD measure may be most affected by the integrity of myelin sheaths while AD may reflect axonal degeneration (Song et al., 2002). Diffusion tensor imaging studies in non-human animals have associated higher RD values with dysmyelination (Tyszka et al., 2006) and demyelination (Harsan et al., 2006; Song et al., 2005). In contrast, lower AD values are thought to reflect degeneration of the axon itself (Sun et al., 2006).

One theory is that chronic pathology in the cellular matrix such as gliosis (Pierpaoli et al., 2001) and a corresponding reduction in extra-cellular water (Niogi and Mukherjee, 2010) may enhance diffusivity along the principle eigenvector and increase AD. A recent report found an increase in AD for the right inferior fronto-occipital fasciculus in veterans with the chronic symptoms commonly referred to as Gulf War Illness (Rayhan et al., 2013). However, others have cautioned against using RD and AD measures to make specific statements regarding the biophysical properties of damaged tissue (Wheeler-Kingshott and Cercignani, 2009). Pathology may affect the direction of the principal eigenvector, which, in turn, artificially increases or decreases RD and AD values. This bias may vary between individual data sets, making any relationship between the measures and pathology suspect.

Like the MD analysis, the stepwise regression did not identify LIF, but RAT, as the best predictor for the RD and AD measures. As discussed above, this may be due to a limited decision tree or inherent differences between the analyses. However, it is notable that LIF is the optimal classifier for three out of the four measures: MD, RD, and AD. A concomitant increase in RD and AD was observed in chronic TBI patients (Kraus et al., 2007). Increases in MD are even more common. We suggest this pattern of results is not coincidental. The combination of higher MD, RD and AD values localized to single tract is likely pathology.

Alternatively, if AD and RD measures suffer from a type of random error, as suggested above, they should not consistently identify damaged tracts. This means further research should not produce the same pattern of DTI measures (i.e., low FA and high MD, RD, and AD values, respectively).
possibly co-located on the same tracts). Accordingly, we suggest continued research with more patients is necessary to address the validity of the RD and AD measures as well as their relevance to the ROC approach.

4.4. Cutpoints

The optimal cutpoints varied considerably between measures. However, the 95% CIs for all cutpoints were relatively small given the range of potential values (Table 2). For example, our analysis found the optimal cutpoint for LCG-FA was 0.433, 95% CI [0.397–0.452]. Fractional anisotropy values have a potential range between 0 and 1. The actual LCG-FA values from our data set ranged between 0.216 and 0.532. Thus, the cutpoint’s confidence interval is approximately 1/20th the potential range, 1/6th the actual range. Narrow CIs indicate our sample supports a precise estimate.

There is currently little evidence regarding what DTI thresholds are diagnostic of TBI. Most DTI research on TBI patients has been cross-sectional in nature. Finding group differences suggests some deviation from the norm is diagnostic. However, between-group comparisons cannot provide a meaningful threshold, only an assurance that the groups do, in fact, differ. An advantage of the exploratory ROC is that it searches each continuous variable for a division that produces the best agreement with the outcome variable. The end result is the optimal, diagnostic threshold. More research with more subjects is necessary, but our findings suggest exploratory ROC may help identify DTI thresholds indicative of TBI pathology.

4.5. Kappa values

When validating medical tests, researchers often look for the degree of consistency between different methods or “raters.” However, comparing the output of any two raters produces a certain amount of agreement simply due to chance. Chance accord inflates the relative number of cases in agreement. This happens regardless of inherent commonalities or differences between the raters. Consequently, a simple percentage will always overestimate the extent of actual agreement (Gwet, 2014).

Inter-rater reliability coefficients take chance agreement into account when comparing raters of categorical data. The kappa coefficient is perhaps the most widely used of these indices (Cohen, 1960). Here the percent of expected agreement (Pe) is used to adjust the percent of observed agreement (Po): \[ \kappa = (Po - Pe) / (1 - Pe). \] Cohen’s kappa is frequently employed in medical research to assess the precision of diagnostic tests (Kraemer et al., 2005). The kappa values reported here comprise some of the first information on the inter-rater agreement between DTI measures and TBI diagnosis.

Cohen’s kappa ranges between 0 (chance) and 1 (perfect agreement). Thus, all of our values are in the lower half of the scale. The benchmark scale from Landis and Koch (1977) regards kappa values between 0.210 and 0.400 as indicative of fair agreement. Applying Landis and Koch’s interpretation, our kappa values and their confidence intervals indicate the optimal classifiers are in fair agreement with TBI diagnosis. However, another scale by Fleiss (1981) considers all values below 0.400 as indicative of poor agreement.

Ultimately, accumulated experience should determine which kappa values are deemed favorable, not predefined scales. The context of our results must be considered. The natural heterogeneity of TBI may oblige lower kappa values for individual tracts. Because of the diffuse nature of axonal injury, it is simply unlikely that any one tract will be highly diagnostic across patients. Our data seem to confirm this reasoning.

While the diagnostic value of any one tract may be limited, this does not negate the purpose of ranking tracts relative to one another. Understanding these relationships could aid in developing algorithms that predict TBI status or outcomes. A fundamental step in algorithm creation is feature engineering. This involves aggregating or decomposing raw data to better represent an underlying problem. Tracts may be uninformative in isolation, but predictive as a composite. Knowing which tracts to combine and how is crucial in developing these models. The decision trees provided by exploratory ROC could aid in this effort.

4.6. ROC and TBI models

Other ROC analyses have assessed relationships between various metrics (demographics, physician observations, neurological exam results, CT findings, etc.) and TBI outcomes. Models such as the International Mission for Prognosis and Analysis of Clinical Trials in TBI (IMPACT) (Maas et al., 2007), Corticosteroid Randomization After Significant Head Injury (CRASH) (MRC CRASH Trial Collaborators, 2008), the Marshall classification system (Marshall et al., 1992), and the Rotterdam CT (Maas et al., 2005) combine several variables to predict TBI mortality (Roozenbeek et al., 2012a, 2012b) and outcomes (Roozenbeek et al., 2012a, 2012b). The sensitivity and specificity of these multi-variable models vary depending on the data set under scrutiny. Area under the curve (AUC) statistics often range between 0.7 and 0.9 (Steyerberg et al., 2008; Lingsma et al., 2013; Honeybul et al., 2014).

Unlike the above models our analysis exclusively used DTI data. Only a handful of ROC models have incorporated DTI. Some have evaluated whole brain measures (Kim et al., 2013); others have examined select tracts (Ressel et al., 2016). A recent paper by Galanaud et al. (2012) combined DTI measures from several tracts to predict TBI outcomes. They found a DTI composite was more predictive of the Glasgow Outcome Scale (GOS) than IMPACT. Our paper differs from Galanaud et al. (2012) in its examination of several tracts rather than a combination. Also, our approach attempts to classify patients based on TBI status rather than outcome. To our knowledge, our study is the first to apply ROC methods to DTI data in order to rank fiber tracts according to their diagnostic potential.

4.7. Sensitivity and specificity

Our ROCs demonstrated high sensitivity (69%–74%), but lower specificity (53%–61%) Table 2. Most clinical tests are not perfectly diagnostic (100% sensitivity, 100% specificity) and must find a reasonable trade-off between sensitivity and specificity that is context dependent. In the case of chronic TBI, an MRI-based test with high sensitivity but reasonable specificity could be useful in corroborating behavioral symptoms. While this concept is intriguing, our goals here were strictly exploratory. We did not seek to validate a medical test, but to explore a method potentially relevant to TBI diagnosis. Further research is necessary to apply these and future results to clinical aims.

It is encouraging, though, that a single tract could attain 74% sensitivity. As with kappa, we consider the possibility of improving accuracy by combining data from multiple tracts into composite measures (Shen, 2008; Su and Liu, 1993). The sensitivity and/or specificity of composites may surpass any of their constituents. In fact, we found that crudely combining data from the LIF-MD and DST-MD increased sensitivity to 89% but reduced specificity to 44%.

More advanced composites may integrate data from several tracts. Multivariate ROC (MultiROC) provides a method for making logical variable combinations (Shultz, 1995). It uses Boolean operators to incorporate two test results. One has a fixed threshold (e.g. Test A > constant); the other is allowed to vary (Test B > variable). The MultiROC analysis generates a ROC curve for the combination (Test A + Test B). Its optimal sensitivity and specificity are then compared to that of the fixed test (Test A). Statistics (AUC) determine whether or not the new variable (Test B) adds diagnostic value.

Running an exploratory ROC beforehand may improve the validity of the MultiROC procedure. Because the ROC determines the most diagnostic tract in the sample, it provides an excellent candidate for the
fixed test. Other tracts, identified further down the decision tree, could serve as varying tests. Wu et al. (2013) recently used a MultiROC technique to assess the interaction of multiple indices in relation to Alzheimer’s disease. For practicality, they limited their analysis to a few measures. It is likely exploratory ROC would be serviceable here and other situations where brute-force analysis is needed to identify relevant variables from a larger set.

4.8. Left hemisphere bias

An interesting aspect of our data is that the tracts with the highest kappa values are all located in the left hemisphere. While this pattern may simply be a peculiarity of this data set, it may also demonstrate an asymmetry in the structure of brain white matter. Klingberg et al. (1999) postulated that left hemisphere axons are more densely packed than the right. This could make them more resistant to trauma and disease. If this is the case, then there could be less variation in DTI measures from the left hemisphere, resulting in a more distinct separation between normal and abnormal. The left hemisphere then may be more diagnostic of pathology simply because its DTI values have a limited range.

To investigate asymmetries in our data we examined 9 tracts with left/right homologs (anterior thalamic, cingulum, cingulum-hippocampus, cortico-spinal, inferior fronto-occipital, inferior longitudinal, superior longitudinal, superior-temporal, and uncinate). We calculated means and standard deviations for each tract across participants. The standard deviation analysis showed no interpretable patterns. However, the mean analysis revealed that 8 out of the 9 tracts have significantly higher FA values in the left hemisphere. For the MD measure, 6 out of 9 tracts have higher values in the right hemisphere. This pattern occurred in the same tracts in both TBI and non-TBI patients. The cingulum and inferior fronto-occipital fasciculus were among them.

Other DTI studies have reported a left hemisphere bias in measures from the cingulum (Gong et al., 2005; Huster et al., 2009; Yin et al., 2013) and inferior fronto-occipital fasciculus (Rodrigo et al., 2007; Thiebaut de Schotten et al., 2011). The etiology of these asymmetries is not fully understood but may exist because certain left hemisphere tracts have higher structural integrity than their right hemisphere homologs. It is possible that the interaction of these natural asymmetries and pathology make certain tracts more diagnostic to the ROC.

Evidence also suggests that aging affects the health of the hemispheres differently. White matter undergoes a gradual degeneration beginning as early as middle age (Yeatman et al., 2014). The right hemi-aging model posits a faster rate of degeneration in right hemisphere tissue leading to more pronounced deficits in visual (right hemisphere) compared to language (left hemisphere) abilities (Brown and Jaffe, 1975). Another model, the hemispheric asymmetry reduction in old adults (HAROLD), argues perceived asymmetries in behavioral performance are actually due to older brains becoming less lateralized in their functioning (Cabeza, 2002). This increase in bilateral processing, however, could ultimately be caused by greater deterioration in the right hemisphere.

Our data do not distinguish between these models or necessarily support an account of asymmetric aging. We discuss these theories only because they offer context for one aspect of our findings. If there is a slight bias in the degeneration of some white matter tracts, this may be exacerbated by the physiological changes associated with chronic TBI. At the moment, the interaction between brain trauma, chronicity, and processes of normal aging is a new but developing area of neuroscience (Smith et al., 2013).

4.9. Limitations

Though most of our findings concur with the TBI/DTI literature, our approach has limitations. Post-traumatic stress disorder is one complication. There is high comorbidity between TBI and PTSD (Bryant, 2001; Hoge et al., 2008; Kennedy et al., 2007; Moore et al., 2006). Patients with TBI commonly experience PTSD symptoms (e.g., hyperarousal, avoidance, depression, anxiety, fatigue, poor concentration, irritability, insomnia) weeks to months after their initial injury (Bombardier et al., 2006; Bryant et al., 2010; Zatzick et al., 2010). The cause of this comorbidity remains unclear. The physical pathology of TBI may precipitate PTSD-like symptoms. However, the stress and emotional intensity associated with a TBI could do the same.

In our sample, the TBI group had 57 (78%) patients with PTSD. The non-TBI group had 21 (58%) patients with PTSD. This discrepancy is not ideal if attempting to control for the effect of PTSD. Accordingly, we cannot be sure that over-representation of PTSD in the TBI group had no effect on the ROC analysis. Future investigations with this method will better counterbalance PTSD status between the groups. However, there is arguably more ecological validity in investigating TBI and PTSD as comorbid conditions than apart.

To better understand the potential influence of PTSD on our findings we re-ran the ROC with PTSD as the outcome variable. The analysis yielded the following tracts as optimal classifiers: LAT for the FA measure, RIL for MD, LAT for RD, and RCG for AD. Logistic regression found RCG was the most predictive tract for the AD measure. No other tracts were verified. Previous DTI research links pathology in the cingulum to the presence of anxiety, depression, and PTSD (Abe et al., 2006; Fani et al., 2012; Isaac et al., 2015; Sanjuan et al., 2013; Sekiguchi et al., 2014). However, there is still a paucity of information on the association between white matter and PTSD (Daniels, 2016).

We acknowledge that PTSD status may have influenced these results. However, our study did not intend to dissociate the effects of TBI and PTSD in white matter. Our goal in this case was to simply evaluate a potential classification method in relation to existing knowledge. Interestingly, ROC analysis with either outcome identifies the cingulum, a tract associated with both TBI and PTSD. These findings suggest that ROC is sensitive to white matter deterioration, but perhaps cannot discriminate TBI and PTSD pathology in the current sample. We suggest further research applying the ROC to larger data sets to identify tract combinations indicative of specific patient groups: those with TBI, those with PTSD, and those with both TBI and PTSD.

Another issue is how to set the initial parameters for the ROC analysis. Here we weighted false positives and false negatives the same (0.5). They contributed equally to calculations of sensitivity and specificity. However, the relative importance of these values may change depending on the clinical context. For patients with established symptoms (e.g., post-concussive syndrome), high sensitivity, at the expense of specificity, may be justified for identifying relevant neurocorrelates. Experimenting with these parameters in different patient groups (acute, sub-acute, and chronic) will likely reinforce the clinical importance of certain tracts and demonstrate how they may change with time.

Final considerations are sample size and the generalizability of the results. While sufficient, our sample size (n = 109) was close to the minimum number needed to run the ROC. For this reason, the analysis automatically stopped after only one division for most measures. Only MD produced a second partition. Including more participants may strengthen the current results and identify other tracts of diagnostic value.

However, even with additional subjects, the generalizability of the results is still an issue. Our study examined adult, former-military TBI patients with chronic symptoms. Results from this sample may not be applicable to cases of acute TBI or chronic TBI in civilians and children. However, it is possible that exploratory ROC could be applied to such samples. Further research with diverse populations is needed for a comprehensive understanding of the relationships between white matter tracts and TBI. We believe the methods presented here could aid in this work.
5. Conclusions

The current study used exploratory ROC analysis to rank fiber tracts according to their agreement with TBI diagnosis. To our knowledge, our study is the first to apply this method in assessing the diagnostic utility of DTI data. The information garnered here could potentially inform future research and aid in hypothesis generation. Extensions of this method may also yield clinically relevant information on the natural history of TBI and its variations.

Techniques in DTI are becoming increasingly influential in the evaluation of brain injury. An understanding of how DTI findings overlap with the conclusions drawn from traditional diagnostics (e.g., clinical screening and neuropsychological tests) is critical to its development. Our study sought to evaluate a new method toward this end. The ROC demonstrates potential for exploring relationships between DTI data and clinical outcomes.

Our results also have relevance to new approaches in precision medicine. Here the goal is to leverage computation toward customized patient care. Classifying disease sub-types is fundamental to this aim. To date, neuroimaging research has failed to resolve differences between certain types of TBI. Perhaps the most critical distinction is why some TBI patients fully recover while others continue to experience symptoms. The method outlined here may be applied to this question and many others. We submit exploratory ROC has the potential to sift through TBI data and generate clinically meaningful classifications. This information may ultimately guide the development of advanced biomarkers for clinical practice.

Acknowledgements

This research was funded by the ar Related Illness and Injury Study Center. F.P. is supported by NSF IIS 1636893 and NIH ULTR001108.

References


