Multi-tract multi-symptom relationships in pediatric concussions

Guido I. Guberman¹, Sonja Stojanovski²,³, Eman Nishat²,³, Alain Ptito¹, Danilo Bzdok⁴,⁵,⁶, Anne Wheeler²,³, Maxime Descoteaux⁷

1. Department of Neurology and Neurosurgery, Faculty of medicine, McGill University, Montreal, Quebec, Canada
2. Department of Physiology, Faculty of Medicine, University of Toronto, Toronto, ON, Canada
3. Neuroscience and Mental Health, The Hospital for Sick Children, Toronto, ON, Canada
4. McConnell Brain Imaging Centre (BIC), Montreal Neurological Institute (MNI), Faculty of Medicine, McGill University, Montreal, Quebec, Canada
5. Department of Biomedical Engineering, Faculty of Medicine, School of Computer Science, McGill University, Montreal, Quebec, Canada
6. Mila - Quebec Artificial Intelligence Institute, Montreal, Quebec, Canada
7. Department of Computer Science, Sherbrooke University, Sherbrooke, QC, Canada

Abstract:
The heterogeneity of white matter damage and symptoms in concussions has been identified as a major obstacle to therapeutic innovation. In contrast, the vast majority of diffusion MRI studies on concussion have traditionally employed group-comparison approaches. Such studies do not consider heterogeneity of damage and symptoms in concussion. To parse concussion heterogeneity, the present study combines diffusion MRI (dMRI) and multivariate statistics to investigate multi-tract multi-symptom relationships. Using dMRI data from a sample of 306 children ages 9 and 10 with a history of concussion from the Adolescent Brain Cognitive Development Study (ABCD study), we built connectomes weighted by classical and emerging diffusion measures. These measures were combined into two informative indices, the first capturing a mixture of patterns suggestive of microstructural complexity, the second representing almost exclusively axonal density. We deployed pattern-learning algorithms to jointly decompose these connectivity features and 19 behavioural measures that capture well-known symptoms of concussions. We found idiosyncratic symptom-specific multi-tract connectivity features, which would not be captured in traditional univariate analyses. Multivariable connectome-symptom correspondences were stronger than all single-tract/single-symptom associations. Multi-tract connectivity features were also expressed equally across different sociodemographic strata and their expression was not accounted for by injury-related variables. In a replication dataset, the expression of multi-tract connectivity features predicted adverse psychiatric outcomes after accounting for other psychopathology-related variables. By defining cross-demographic multi-tract multi-symptom relationships to parse concussion heterogeneity, the present study can pave the way for the development of improved stratification strategies that may contribute to the success of future clinical trials and the improvement of concussion management.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Introduction

Concussion afflicts approximately 600 per 100,000 individuals every year.\(^1\) Its incidence rate is rising in children and adolescents,\(^2\) and compared to adult populations, the impact of concussions on pediatric brains is understudied.\(^3\) Despite considerable funding devoted to clinical and basic research, no major advances in therapeutics have been achieved to date.\(^4\) A root cause of this stagnation appears to be a contradiction: while all concussions are treated equally in clinical trials and research studies, they are characterized by extensive heterogeneity in their pathophysiology, clinical presentation, symptom severity and duration.\(^4,5\) Concussion heterogeneity across patients has been identified as a major hurdle in advancing concussion care.\(^4,5\)

Due to shearing forces transmitted during injury, the brain’s white matter is especially vulnerable to concussion.\(^6,7\) Decades of research have studied white matter structure in individuals who sustain concussions. However, most studies continue to assume consistent, one-to-one structure/symptom relationships and employ traditional group comparisons,\(^8,9\) averaging out the diffuse and likely more idiosyncratic patterns of brain structure abnormalities in favour of shared ones. Hence, the extant literature suggests that a large proportion of the clinical and research studies have not adequately accounted for clinical and neuropathological concussion heterogeneity.

To remedy this shortcoming, a growing number of studies aimed to parse the clinical heterogeneity in concussions by algorithmically partitioning patients into discrete subgroups based on symptoms.\(^10-12\) Other studies aim instead to account for heterogeneity in white matter structure alterations.\(^13-15\) Ware et al.\(^15\) built individualized maps of white matter abnormalities which revealed substantial inter-subject variability in traumatic axonal injury and minimal consistency of subject-level effects. Taylor et al.\(^14\) computed a multivariate summary measure of white matter structure across 22 major white matter bundles which achieved better classification accuracy of concussed patients from healthy controls compared to single tract measures. Hence, studies have attempted to address concussion heterogeneity in symptoms and in white matter structure. However, no prior studies have considered both sources of heterogeneity simultaneously.

White matter alterations due to concussion are diffuse and can elicit several symptoms that may interact with each other in complex ways.\(^4,5,16\) For instance, two individuals may suffer a concussion and develop sleep problems. The first may have damaged white matter tracts related to sleep/wakefulness control, whereas the second may have damaged tracts related to mood, causing depression-like symptoms, which include sleep problems. These two individuals will thus display a common symptom but will have overall different symptom profiles and different white matter damage profiles. Parsing concussion heterogeneity requires accounting for these dynamic, multi-tract multi-symptom relationships.

In the present study, we leveraged advanced diffusion MRI (dMRI) methods as well as a double-multivariate approach to parse concussion heterogeneity in white matter structure and symptoms simultaneously in a large sample of previously-concussed children. Multi-tract multi-
symptom relationships captured more information than traditional univariate approaches. Expression of multi-tract connectivity features was not driven by sociodemographic strata and injury-related variables. Finally, after accounting for univariate variables found to be related to adverse psychiatric outcome in the discovery dataset (n=214), we found that expression of one multi-tract connectivity feature predicted adverse psychiatric outcome in a replication dataset (n=92).

Methods

Participants

Participants in this study were obtained from the world’s largest child development study of its kind — the ongoing longitudinal Adolescent Brain Cognitive Development Study (ABCD study; https://abcdstudy.org/), data release 2.0 (https://data-archive.nimh.nih.gov/abcd). The ABCD Study acquired data from 11,874 children aged 9 to 10 years (mean age = 9.49 years) from across the United States (48% girls; 57% Caucasian, 15% African American, 20% Hispanic, 8% other). Additional information about the ABCD study can be found in Garavan et al.18

History of concussion

Parents completed a modified version of the Ohio State University TBI Identification Method (OSU-TBI-ID)19. We included participants who reported a head injury without loss of consciousness but with memory loss and/or a head injury with loss of consciousness for less than 30 minutes (n=434). Due to missing or incomplete data, corrupted files, data conversion errors, and images rated by the ABCD study team as being of poor quality, the final sample of participants with usable data was 345. After processing, images were visually inspected by two independent raters (G.I.G., S.S.). Images that were deemed of low quality after processing by both raters were removed (n=39), leading to a final sample of 306 participants. We randomly divided the sample into a discovery dataset (70%, n=214) and a replication dataset (20%, n=92). Figure 1 summarizes the subject selection procedure.

Symptom-oriented measures

To probe various aspects of concussion symptomatology, we used items collected from assessments available in the ABCD dataset. These items, as well as the concussion symptom they are meant to probe are outlined in Table 1.

MRI Acquisition

MRI scans were acquired across 21 sites, with data coming from 28 different scanners. Details about the acquisition protocols and image specifications are outlined in Casey et al 2018. Multi-shell dMRI scans had 96 diffusion-weighted directions, with 6 directions of b=500 s/mm², 15 directions of b=1000s/mm², 15 directions of b=2000s/mm², and 60 directions of b=3000s/mm². The b=2000 shell was excluded from the data processing. In addition, scans had 6 or 7 b=0s/mm² images, depending on scanner type. Lastly, a reverse b0 image was included for each participant.

Processing
We used Tractoflow\textsuperscript{21} to process dMRI and T1-weighted scans. Tractoflow is a novel diffusion MRI processing pipeline, incorporating state-of-the-art functions from FSL, Dipy, and MRtrix into NextFlow. The processing steps are summarized in Theaud et al.\textsuperscript{21} Important deviations from the default parameters utilized by Tractoflow are as follows: 1. We used gray-white matter interface seeding, as this method accounts for the length bias introduced by white-matter seeding;\textsuperscript{22} 2. We used 24 seeds-per-voxel with the objective of obtaining approximately 2 million streamlines across the entire brain. We used the b=0, 500, and 1000 shells to perform tensor fitting, and the b=0 and 3000 shells to perform Constrained Spherical Deconvolution (CSD) (Descoteaux et al. 2009; Tournier et al. 2007).\textsuperscript{23,24} We obtained group-average fiber-response functions from voxels with high (>0.70) fractional anisotropy (FA). Lastly, we created tractograms using a probabilistic particle-filtering tractography algorithm (Girard et al. 2014).\textsuperscript{22}

Connectivity matrices
The post-processing workflow is illustrated in Figure 2. To construct connectivity matrices, we used Freesurfer on McGill's CBrain platform\textsuperscript{25} to fit the Desikan-Killiani Tourville (DKT)\textsuperscript{26} and aseg atlases onto the processed T1-images that had been transformed to DWI space during processing (Figure 2A). We removed redundant and irrelevant labels from the fitted atlas (a list of retained labels is supplied in supplementary material), yielding a final atlas with 76 labels. We then thresholded matrices such that a connection was only retained if it was found to be successfully reconstructed (defined as the presence of at least one streamline) across 90% of participants.\textsuperscript{27} Results using different thresholds are presented in supplementary material. We then weighted thresholded connectomes by FA, mean, radial, and axial diffusivities (MD, RD, AD respectively), number of fiber orientations (NuFO), and apparent fiber density along fixels (AFDf) (Figure 2B). The first four measures are derived from the tensor model, whereas the latter two are based on fiber orientation distribution functions (fODFs) obtained from CSD.\textsuperscript{28,29} Simulation studies have shown that AFD is more specifically related to axonal density, and by computing it along “fixels” (fiber elements), axonal density specific to particular fiber populations can be studied independently of crossing fibers.\textsuperscript{28} Although individual diffusion measures are related to different aspects of neuropathology, together they provide more information than when considered separately.\textsuperscript{27} A recent framework based on principal component analysis (PCA) has been proposed to recombine diffusion measures into biologically-interpretable measures of white matter structure.\textsuperscript{30} We therefore performed PCA on the concatenated set of standardized measures across subjects and connections, generating connectivity matrices weighted by principal component (PC) scores (Figure 2C).

Additional data transformations
We imputed missing connectivity (prior to the PCA), symptom, and nuisance data (sex, pubertal stage, handedness, scanner) by randomly selecting non-missing data from other participants in the same dataset. We reverse-coded cognitive scores, such that increasing scores in all symptom data reflected more problems. From connectivity and symptom data, we regressed out the following nuisance variables: sex, pubertal stage, scanner (only for connectivity data), and handedness.
Pattern-learning pipeline

**Feature selection.** To reduce the number of connectivity features included in the partial least squares correlation (PLSc) analysis, we selected the 200 connectivity features most correlated (based on Pearson correlations) with any symptom score. This solution is becoming increasingly adopted for high-dimensional variable sets (Figure 2D).\(^{31,32}\)

**PLSc.** We performed PLSc analyses using the `tepPLS` function from the `texposition` package.\(^33\) PLSc involves singular value decomposition on the covariance matrix between connectivity and symptom features, creating pairs of multi-tract and multi-symptom features called multi-tract multi-symptom relationships. Each multi-tract multi-symptom relationship encapsulates a linear combination of connectivity features ("multi-tract features"), a linear combination of symptom scores ("multi-symptom features"), and an eigenvalue (reflective of the amount of explained covariance between connectivity and symptom features). Each multi-tract multi-symptom relationship is constructed so as to explain a successively smaller portion of the covariance between symptoms and connectivity features. We constructed the largest number of possible multi-tract multi-symptom relationships, given the dimensionality of the behavioral variable set (k=19) (Figure 2D).

**Selection and interpretation of multi-tract multi-symptom pairs.** To select the multi-tract multi-symptom pairs to retain for interpretation, we performed permutation testing (2000 iterations). This procedure randomly shuffles row labels for the connectivity features, without replacement, repeats the PLSc and computes eigenvalues at every permutation. We calculated \(p\)-values as the proportion of permutations that yielded eigenvalues that exceeded the original amount.

To interpret symptom and connectivity weights of significant \((p<0.05)\) multi-tract multi-symptom pairs, we performed bootstrap analyses (2000 iterations), using the `BOOT4PLSC` command from the `texposition` package. At each iteration, labels for data were drawn with replacement, the entire PLSc was repeated and the weights for all pairs were obtained. This process yields a sampling distribution of weights for each connectivity and symptom feature.\(^{34}\) The ratio of the original weights to the standard error of each feature’s bootstrap distribution can be interpreted as a z-score, which yielded so-called ‘bootstrap ratios’. We used a value of 1.96 to determine which variables significantly contributed to each particular significant pair.

**Comparison of multivariate against univariate approaches**

To compare information captured by the PLSc and univariate approaches, we first divided participants based on whether they had T-scores above 70, a well-recognized clinical threshold,\(^35\) in the Child Behavior Checklist (CBCL) Depression, Attention Problems, Anxiety Disorder, or Aggression scales.\(^35\) Using a threshold of \(p<0.05\), we computed univariate comparisons of connectivity (PC scores) between individuals with and those without clinical-level psychopathology, thus identifying psychopathology-related univariate connectivity features.

We were interested in comparing how many of these features were also found to significantly contribute to each multi-tract connectivity feature. To do so, we computed a measure of percent overlap as follows:

\[
\text{\%Overlap} = \frac{c_{\text{sig}}}{(S_t + S_m - c_{\text{sig}})} \times 100,
\]
where $C_{\text{sig}}$ refers to the number of connections flagged as significant in both approaches, $S_u$ to the number of connections flagged as significant in the univariate approach, and $S_m$ to the number of connections flagged as significant in the multivariate approach. This measure can account for the apparent high overlap that can arise when $S_u$ and $S_m$ are not equivalent in size.

**Relation to TBI-related and sociodemographic factors**

We addressed whether expression of multi-tract connectivity features was related to injury-specific and sociodemographic factors. Injury-related variables included: the time between the last-documented injury and testing, the cause of injury, and the total number of documented mTBIs. Sociodemographic variables included: sex, total combined family income in the last 12 months, and race/ethnicity. We used the following categories for race/ethnicity: “Asian” (Asian Indian, Chinese, Filipino, Japanese, Korean, Vietnamese, Other Asian), AIAN (“American Indian”/Native American, Alaska Native), NHPI (Native Hawaiian, Guamanian, Samoan, Other Pacific Islander), Non-Hispanic White, Non-Hispanic Black, Hispanic, Other, and Multiple. To illustrate the influence of these sociodemographic factors we created scatter plots illustrating expression of connectivity latent factors color-coded by sociodemographic factors (Supplementary Figure 1).

**Prediction of clinical outcome**

We used the Kiddie-Schedule for Affective and Psychiatric Disorders in School Age Children (KSADS), a gold-standard tool, to assess the presence of pediatric psychiatric disorders. We used the presence of any current psychiatric diagnosis as indicating an adverse psychiatric outcome (55% of discovery set, 55% of replication set). We created separate multivariable logistic regressions to predict adverse psychiatric outcomes using 1) expression of multi-tract connectivity features, 2) psychopathology measures from the CBCL, or 3) psychopathology-related univariate connectivity features. From each set of models, we retained the variables that were significantly related to adverse psychiatric outcome. We then tested whether expression of multi-tract connectivity features could predict adverse psychiatric outcome in the replication dataset.

To do so, we first projected connectivity data from the replication dataset onto the same principal components retained in the discovery dataset. Next, we selected the same 200 connections that had been retained based on univariate feature selection, and projected these connections onto the multi-tract connectivity spaces obtained from the discovery dataset. Finally, we performed a logistic regression model incorporating all the predictors found to be significant in the discovery dataset, applied on data from the replication dataset.

The objective of these models was to assess whether expression of multi-tract connectivity features could predict adverse psychiatric outcomes in the replication dataset after accounting for other psychopathology-related factors (CBCL behavioural measures, univariate connectivity measures).

**RESULTS**

*Sample*
Out of 434 participants with a history of mTBI, 306 (127F/179M) had usable data (Figure 1). Table 2 outlines sociodemographic and injury-related factors, as well as handedness and sex. The majority had sustained an injury over 1 year prior to the study. Nuisance variables were well-balanced between participants in the discovery and the replication set.

Recombined measures of white matter tract microstructure

The PCA yielded two biologically-interpretable components that together explained 97% of the variance in dMRI measures (Figure S2). The first (PC1) represented all diffusion measures except AFDf, and likely reflects microstructural complexity. The second (PC2) represented AFDf exclusively, and likely reflects axonal density. Because we retained two PCs, we performed two PLSc analyses.

Multi-tract multi-symptom relationships

Each PLSc analysis yielded 19 latent modes of covariance (termed here “multi-tract multi-symptom relationships”), each consisting of a pair of multi-tract connectivity and multi-symptom features. Based on permutation testing, 18 multi-tract multi-symptom pairs were retained from PC1 and 14 from PC2. For brevity, only a few informative pairs are presented herein, but an illustration of all significant pairs for all thresholds can be found in supplementary material (Figures S3-S6).

For PC1, similarly to PC2, the first multi-tract multi-symptom pair broadly represented all symptoms (Figure 3A), capturing general problems. Interestingly, this pair implicated several callosal tracts (Figure 5). The third multi-tract multi-symptom pair obtained from PC2 represented mostly sleep and cognitive problems, whereas the fourth pair represented more strongly mood, sleep, and somatic problems (Figure 3B and C).

Multivariate vs univariate approaches

Correlations between multi-tract and multi-symptom features were higher than all univariate correlations between the PC1 and PC2 scores of every retained tract and every symptom (“single-tract single-symptom relationships”) (Figure 4).

We identified 28 individuals with scores above 70 in CBCL Depression, Attention Problems, Anxiety Disorder, or Aggression scales, considered to be the clinical range. Figure 3 illustrates the expression of multi-symptom and multi-tract connectivity features from three pairs. In clinical trials and research studies, these 28 individuals would be grouped together. However, when considering all symptoms, these individuals appear to be markedly different. Accordingly, these individuals were differentiable by their expression of multi-symptom and multi-tract connectivity features (Figure 3).

We compared PC1 and PC2 scores across all 200 connections between individuals with and without clinical-level CBCL scores, and calculated the percent overlap between each multi-tract connectivity feature and the set of tracts found to be significant in univariate comparisons. The percent overlap scores are presented in Figure 5. Notably, the highest overlap occurred with multi-tract connectivity feature 1 (22%) from both PLSc analyses (Figure 5). Univariate psychopathology-related features (Figure 5 red and blue chord plots) and multi-tract connectivity feature 1 (Figure 5 violet chord plot) implicated several callosal tracks, whereas
fewer callosal tracks were implicated in multi-tract connectivity features 3 and 4 (Figure 5 turquoise and green chord plots).

Relationship with sociodemographic and injury-related factors

The expression of multi-tract connectivity features was equivalent across sociodemographic strata defined by sex, total combined household income, and race/ethnicity (Supplementary Figure 1). Out of 32 retained multi-tract multi-symptom pairs, time-since-injury was only significantly correlated to the expression of two multi-tract connectivity features (Table S1) and no multi-symptom features. Only one multi-tract connectivity feature (feature 2 from the first PLSc) was significantly different between groups defined by injury cause (Figure S7). Only two multi-tract connectivity features, feature 17 from the first PLSc (Figure S8), and feature 8 from the second PLSc (Figure S9).

Prediction of clinical outcome

Using separate multivariable logistic regressions, we found expression of four multi-tract connectivity features from PC1, one behavioural measure (CBCL Attention Problems), and no single tract features that were significantly related to adverse psychiatric outcome. No other variables, including no multi-tract connectivity features obtained from PC2 scores, were related to adverse psychiatric outcome. We projected the data from the replication set onto the PCs obtained from the discovery set, selected the same 200 connections, and computed the expression of the multi-tract connectivity features significantly related to adverse psychiatric outcome in the discovery set. We incorporated these multi-tract connectivity features with other significant variables, using the replication data, in a single multivariable logistic regression. We found that a one-unit increase in the expression of multi-tract connectivity feature 4 significantly increased the odds of an adverse psychiatric outcome in the replication set 2.31 times (p=0.02) after controlling for univariate, psychopathology-related behavioural features. This multi-tract connectivity feature was paired with a multi-symptom feature that implicated sleep and somatic problems (Figure 6).

Sensitivity analyses

Results using different thresholds (t=85%, 95%, 100%) for the connectomes are outlined in supplementary material. All results were consistent, including PCs, the weights of the multi-symptom features, the strength of multivariate vs univariate relationships, the trends for the % overlap scores, and the significance of adverse psychiatric outcome predictions using multi-tract connectivity feature expression. Results were less consistent with the t=100% threshold, which demonstrated instead little to no significant multi-tract/multi-symptom pairs based on permutation testing. Correlations between expression of multi-tract connectivity features at different thresholds are illustrated in supplementary material. Notably, for t=85%, 90%, and 95%, correlations between the expression of the corresponding multi-tract connectivity features (e.g.: multi-tract feature 1 from t=85%, multi-tract feature 1 from t=95%) all have high correlations, indicating how similar these features are, whereas these correlations were lower when comparing against t=100% (Figure S10).

Discussion
In the present study we leveraged novel dMRI methods and a double-multivariate approach to parse heterogeneity in white matter structure and symptoms in a large sample of previously-concussed children. Applying PLSc on biologically-interpretable measures of dMRI obtained from PCA, we found cross-demographic latent multi-tract multi-symptom relationships that captured more information than traditional approaches and predicted meaningful clinical outcomes in unseen data. These results both recapitulated well-known findings from the concussion literature and revealed new insights about white matter structure/symptom relationships.

In response to decades of failed attempts to translate basic science findings into successful clinical trials and novel therapies, concussion heterogeneity has been identified as a major obstacle.\textsuperscript{4,5} Heterogeneity in symptoms, impact of injury on brain structure and function, and pre-injury factors pose a particular problem for most concussion neuroimaging studies which have traditionally employed univariate comparisons between concussed and healthy or orthopedic injury control groups, or between patients with and without persistent symptoms.\textsuperscript{8,9} These sources of variability are believed to be problematic because they decrease the statistical power needed for group comparisons and multivariable models to detect the often-subtle effects of concussions.\textsuperscript{38} To overcome this challenge, landmark initiatives such as the IMPACT,\textsuperscript{38} InTBIR,\textsuperscript{39} CENTER TBI,\textsuperscript{40} and TRACK TBI\textsuperscript{41} aim to standardize and pool multi-center data collected across sociodemographic strata, to identify and statistically correct for pre-injury factors known to impact brain structure, and develop diagnostic and prognostic tools leveraging multimodal data and increasingly sophisticated machine-learning approaches.

In this study, we posited that concussion heterogeneity is also problematic because by pooling across patients, idiosyncratic patterns of connectivity that may be more symptom-specific are sacrificed in favour of shared ones. By assuming that symptoms map cleanly and consistently onto shared connectivity abnormalities in a one-to-one fashion, erroneous inferences could be made about relationships between group-level patterns of connectivity differences and specific symptoms. Our results are consistent with this idea: univariate comparisons between a group displaying clinical-level psychopathology and the rest of the sample identified connectivity features that mostly overlapped with the first multi-tract multi-symptom pair obtained from both PLSc analyses. These pairs, which accounted for the most covariance, reflected general problems, not specifically psychopathology. Both these pairs and the univariate “psychopathology-related” connections implicated several callosal tracts. These results suggest that univariate comparisons, even when performed in such a way as to identify a symptom-specific set of connectivity features, identified only the most consistent group-level connectivity differences at the expense of more symptom-specific idiosyncratic ones. These findings are consistent with most prior concussion literature, which strongly implicate the corpus callosum,\textsuperscript{9} but extend this literature by suggesting that the central importance of the corpus callosum may have resulted from pooling across concussed subjects.

Rather than a clean and consistent set of single-tract single-symptom relationships, this study suggests concussions may be best conceptualized as a multiplicity of multi-tract multi-symptom combinations. Multi-tract relationships may be driven by the metabolic demands...
imposed by the network structure of the brain, which is known to predict the course of several brain diseases,\textsuperscript{42} by biomechanical constraints imposed by the skull and other structures exposing certain areas to more shearing strain,\textsuperscript{43} or by both factors simultaneously.\textsuperscript{44} These possibilities need to be tested further.

Few recent studies have parsed heterogeneity in concussion symptoms using clustering analyses.\textsuperscript{10-12} The reported subgroups differed from those found in the present study. Differences between symptom profiles arose because our multi-symptom features are associated with brain structure and not driven by variability in symptoms alone. Further, we did not group patients into discrete subgroups. Although whether patients cleanly fit into discrete subtypes has not been explicitly tested in our study, expression of multi-tract and multi-symptom features do not suggest any obvious clusters. The existence of concussion subtypes needs to be explored further.

Prior studies have attempted to address heterogeneity in white matter structure in concussions.\textsuperscript{13-15} Using different approaches, these studies generated point summaries that accounted for the high-dimensional variability of white matter structure to better distinguish patients from controls. Our approach offers an important advantage. Rather than collapsing the highly variable information provided by white matter structure, our approach attempts to find symptom-specific patterns of white matter structure, which has potential to lead to new treatment targets.

After accounting for behavioural and single-tract predictors of adverse psychiatric outcome, the expression of one multi-tract connectivity feature significantly predicted adverse psychiatric outcome in a holdout dataset. The multi-symptom feature that corresponded to this significant multi-tract feature implicated mostly sleep and somatic problems. Sleep problems are known to be linked to psychiatric disorders, both as symptoms, and risk factors.\textsuperscript{45} Longitudinal studies are needed to elucidate whether these multi-tract features reflect abnormalities that increase the risk of later adverse psychiatric outcomes, or instead reflect abnormalities associated with existing psychiatric diagnoses. Nonetheless, the prediction of a meaningful clinical outcome on unseen data is particularly promising.

\textit{Limitations and Strengths}

The present results should be considered in light of methodological limitations. Data on mTBI occurrence was collected retrospectively. Participants did not have baseline data, and additionally had highly variable times since injury. Most individuals with concussions recover from their injury\textsuperscript{46} which should have led to a concussed group where most participants were similar to healthy controls. Interestingly, our PCA yielded combinations of diffusion measures that differed from those of two prior studies that have used this approach.\textsuperscript{30,47} These prior studies used samples of typically-developing children without neurological insults and found similar PCs. To the extent that the PCs reported in these prior studies reflect healthy neurotypical brains, our PCs suggest that our concussed sample was not as similar in white matter structure to healthy controls as expected. However, variable time since injury made the interpretation of patterns of microstructure difficult. Further, due to the cross-sectional nature
of this data, the difference between symptoms and pre-existing characteristics are difficult to
discern, especially since some behavioural measures often believed to be symptoms of
concussion, such as attention problems, can also be risk factors for injury.\textsuperscript{48} Heterogeneity has
several forms, including in symptoms, duration, severity, neuropathology,\textsuperscript{7} lesion location,\textsuperscript{15}
sociodemographics,\textsuperscript{38} genetics,\textsuperscript{49} behaviour,\textsuperscript{48} pre-injury comorbidities,\textsuperscript{50} and environmental
differences, including access to and quality of care.\textsuperscript{51} These factors have been theorized to
interact in complex ways.\textsuperscript{4} This study only addressed a minority of these complex relationships,
further studies integrating even more sets of data are needed to address these other drivers of
heterogeneity.

Conversely, this study leveraged some of the most recent and important advances in
dMRI to address the major limitations of conventional approaches. We used high quality multi-
shell dMRI data,\textsuperscript{52} as well as modelling approaches, tractography techniques, and
microstructural measures robust to crossing fibers, partial volume effects, and connectivity
biases.\textsuperscript{22-24,28-30} We used PCA to combine dMRI measures into meaningful indices of white
matter structure. Lastly, we used gold-standard measures of psychiatric illness to predict clinical
outcomes. Future iterations of this work will need better control of time since injury,
longitudinal follow-up, broader assessment of clinical outcomes, and further development of
predictive models.

Conclusion

In conclusion, leveraging advanced dMRI and a pattern-learning algorithm to parse
concussion heterogeneity, we have found clinically-meaningful cross-demographic latent
representations of multi-tract multi-symptom relationships. As the field moves towards large-
scale studies aiming to statistically control for sociodemographic sources of heterogeneity to
detect a putative consistent white matter signature of concussion across patients, the insights
gained from this study should be taken into consideration: informative, clinically-meaningful,
symptom-specific patterns of connectivity differences may be lost when pooling across
concussed patients. This insight is an important step in improving stratification strategies for
clinical trials and the identification of treatment targets.
References


Figure 1. Flowchart describing the participant selection procedure.
Figure 2. Schematic illustration of the study’s workflow. A. After processing DWI images using tractoflow, we applied the DKT parcellation onto each participant’s T1-weighted image that had been registered to DWI space. From each tractogram, we extracted streamlines connecting each pair of labels in the DKT parcellation. Using this approach, we built a binary connectivity matrix that displayed for all 306 subjects in the full dataset (rows), whether (black) or not (white) a streamline existed between each pair of labels (columns). B. We split the dataset into a discovery set (n=214) and a replication set (n=92). On the discovery set, we thresholded connectomes, only keeping connections that existed across 90% of participants (a threshold of 100% is illustrated here for simplicity). We then constructed connectomes of 6 scalar diffusion measures (Fractional Anisotropy (FA), Axial Diffusivity (AD), Mean Diffusivity (MD), Radial Diffusivity (RD), Apparent Fiber Density along fixels (AFDf), and Number of Fiber Orientations (NuFO)), by computing the average measures across each bundle (weighted by the streamline count across each voxel to reduce the impact of spurious streamlines). C. We stacked all columns from each connectivity matrix, creating vectors every pair of subject/connection features. We then joined together these vectors into one matrix, showing pairs of subject/connection features (rows) for every diffusion measure (columns). We then performed principal component analysis (PCA) on these matrices. Given than each pair of subject/connection feature was treated as an observation, principal component (PC) scores could be calculated for each subject/connection feature, allowing us to reconstruct connectomes, weighted by PC scores rather than the original diffusion measures. D. From these connectomes of PC scores, we selected 200 features on the basis of univariate Pearson correlations with symptom-oriented measures. We then performed partial least squares correlation on these PC-weighted features and symptom measures, which allowed us to obtain multi-tract/multi-symptom relationships, pairs of multi-tract connectivity features (“MCF”) and multi-symptom features (“MSF”). Each multi-variate feature is composed of linear combinations (weighted sums, illustrated by the black arrows called “weights”) of variables from its corresponding feature set.
Figure 3. Illustration of 3 multi-tract multi-symptom relationships obtained from PLSc on selected PC1 features (A), or selected PC2 features (B and C). Left column: Polar plots displaying the weights of all 19 symptom-oriented measures on multi-symptom features. Each circle is color-coded to match with the corresponding scatter plots. Bars pointing away from the circle center illustrate positive weights, bars pointing towards the circle center represent negative weights. White stars illustrate symptoms that were found to significantly contribute to the multi-tract multi-symptom pair based on bootstrapping analyses. The bar graph underneath the polar plots illustrates the percent covariance explained by each multi-tract multi-symptom pair, with the currently-shown pair highlighted. Right column: Scatter plots showing the expression of multi-tract connectivity features (x-axis) and multi-symptom features (y-axis). In each plot, the same 6 representative participants are labelled (1 through 6), with each plot showing the scaled symptom-oriented measures (i.e.: not the expression of multi-symptom features) for two participants, one expressing low levels of a multi-tract multi-symptom pair, the other expressing high levels of the pair. For each illustrated participant, positive bars illustrate symptoms that are higher than the sample average, negative bars represent symptoms that are lower than the sample average. Participants with clinical levels of CBCL Depression, Attention Problems, Anxiety Disorder, or Aggression scores are illustrated in black. A. Illustration of multi-tract multi-symptom pair 1 obtained from the PLSc performed on selected PC1 features. B. Illustration of multi-tract multi-symptom pair 3 obtained from the PLSc performed on selected PC2 features. C. Illustration of multi-tract multi-symptom pair 4 obtained from the PLSc performed on selected PC2 features.
Figure 4. Histogram illustrating correlation coefficients for all possible single-tract single-symptom relationships (blue), and between corresponding connectivity and symptom latent factors (pink) (referred to as “multi-tract multi-symptom relationships”).
Figure 5. Line plot showing the percent overlap between univariate analyses and each connectivity latent factor. Highest overlap occurred for the LF1 from both PLSc analyses. Chord plots shown above graph illustrate which connections were found to be significant for univariate comparisons of PC1 features (red), univariate comparisons of PC2 features (blue), multi-tract connectivity feature 1 from the PLSc performed on PC1 features (violet), multi-tract connectivity feature 3 from the PLSc performed on PC2 features (turquoise), and multi-tract connectivity feature 4 from the PLSc performed on PC2 features (green). Inter-hemispheric connections are identified in each chord plot by a black outline, illustrating how many interhemispheric connections are present for univariate comparisons and for multi-tract connectivity feature 1, but less so for multi-tract connectivity features 3 and 4. The percent overlap score for each of the three illustrated multi-tract connectivity features are identified in the line plot with a circle of the corresponding color.
Figure 6. A. Bar plot illustrating the average expression of multi-tract connectivity feature 4 for participants with and without adverse psychiatric outcomes. B. Polar plot illustrating weights of individual symptoms contributing to multi-symptom feature 4.
Table 1.

<table>
<thead>
<tr>
<th>Questionnaire - Description</th>
<th>Symptom Measured</th>
<th>Respondent</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBCL – Headaches</td>
<td>Headaches</td>
<td>Parent</td>
</tr>
<tr>
<td>CBCL – Nausea, feels sick</td>
<td>Nausea</td>
<td>Parent</td>
</tr>
<tr>
<td>CBCL – Vomiting, throwing up</td>
<td>Vomiting</td>
<td>Parent</td>
</tr>
<tr>
<td>CBCL – Feels dizzy or lightheaded</td>
<td>Dizziness</td>
<td>Parent</td>
</tr>
<tr>
<td>CBCL – Overtired without good reason</td>
<td>Fatigue</td>
<td>Parent</td>
</tr>
<tr>
<td>SD – Sleeps less than most kids</td>
<td>Sleep more than usual</td>
<td>Parent</td>
</tr>
<tr>
<td>CBCL – Depression (DSM) T score</td>
<td>Sadness</td>
<td>Parent</td>
</tr>
<tr>
<td>CBCL – Anxiety Disorder (DSM) T score</td>
<td>Nervousness</td>
<td>Parent</td>
</tr>
<tr>
<td>CBCL – Attention Problems T score</td>
<td>Difficulty concentrating</td>
<td>Parent</td>
</tr>
<tr>
<td>NIH Toolbox Picture Sequence Memory Test – Fully-Corrected T-score</td>
<td>Sequence Memory (difficulty remembering)</td>
<td>Child</td>
</tr>
<tr>
<td>NIH Toolbox List Sorting Working Memory Test – Fully-Corrected T-score</td>
<td>Working memory (difficulty remembering)</td>
<td>Child</td>
</tr>
<tr>
<td>RAVLT Short Delay Trial VI – Total Correct</td>
<td>Short recall (difficulty remembering)</td>
<td>Child</td>
</tr>
<tr>
<td>RAVLT Long Delay Trial VII – Total Correct</td>
<td>Long recall (difficulty remembering)</td>
<td>Child</td>
</tr>
<tr>
<td>CBCL Aggression T score</td>
<td>Irritability</td>
<td>Child</td>
</tr>
</tbody>
</table>

Table outlining all behavioural measures used in analyses, along with the corresponding symptom they reflect.

<table>
<thead>
<tr>
<th>Demographic and injury data</th>
<th>Discovery set (n=214)</th>
<th>Replication set (n=92)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Interview Age</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>9.57 (0.496)</td>
<td>9.54 (0.501)</td>
</tr>
<tr>
<td>Median [Min, Max]</td>
<td>10.0 [9.00, 10.00]</td>
<td>10.0 [9.00, 10.0]</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>88 (41.1%)</td>
<td>39 (42.4%)</td>
</tr>
<tr>
<td>M</td>
<td>126 (58.9%)</td>
<td>53 (57.6%)</td>
</tr>
<tr>
<td><strong>Pubertal Stage</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Early</td>
<td>41 (19.2%)</td>
<td>18 (19.6%)</td>
</tr>
<tr>
<td>Mid</td>
<td>58 (27.1%)</td>
<td>19 (20.7%)</td>
</tr>
<tr>
<td>Prepubertal</td>
<td>115 (53.7%)</td>
<td>52 (56.5%)</td>
</tr>
<tr>
<td>Late</td>
<td>0 (0%)</td>
<td>3 (3.3%)</td>
</tr>
<tr>
<td><strong>Race/Ethnicity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>2 (0.9%)</td>
<td>2 (2.2%)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>27 (12.6%)</td>
<td>18 (19.6%)</td>
</tr>
<tr>
<td>Multiple</td>
<td>18 (8.4%)</td>
<td>8 (8.7%)</td>
</tr>
<tr>
<td>Non-Hispanic Black</td>
<td>14 (6.5%)</td>
<td>11 (12.0%)</td>
</tr>
<tr>
<td>Non-Hispanic White</td>
<td>151 (70.6%)</td>
<td>52 (56.5%)</td>
</tr>
<tr>
<td>Other</td>
<td>2 (0.9%)</td>
<td>1 (1.1%)</td>
</tr>
<tr>
<td><strong>Combined Family Income</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;5K</td>
<td>5 (2.3%)</td>
<td>5 (5.4%)</td>
</tr>
<tr>
<td>$5,000 - $11,999</td>
<td>5 (2.3%)</td>
<td>1 (1.1%)</td>
</tr>
<tr>
<td>$12,000-$15,999</td>
<td>3 (1.4%)</td>
<td>2 (2.2%)</td>
</tr>
<tr>
<td>$16,000-$24,999</td>
<td>5 (2.3%)</td>
<td>3 (3.3%)</td>
</tr>
<tr>
<td>$25,000-$34,999</td>
<td>12 (5.6%)</td>
<td>4 (4.3%)</td>
</tr>
<tr>
<td>$35,000-$49,999</td>
<td>12 (5.6%)</td>
<td>5 (5.4%)</td>
</tr>
<tr>
<td>$50,000-$74,999</td>
<td>34 (15.9%)</td>
<td>16 (17.4%)</td>
</tr>
<tr>
<td>$75,000-$99,999</td>
<td>31 (14.5%)</td>
<td>13 (14.1%)</td>
</tr>
<tr>
<td>$100,000-$199,000</td>
<td>76 (35.5%)</td>
<td>27 (29.3%)</td>
</tr>
<tr>
<td>&gt;$200,000</td>
<td>31 (14.5%)</td>
<td>16 (17.4%)</td>
</tr>
<tr>
<td><strong>Handedness</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LH</td>
<td>10 (4.7%)</td>
<td>10 (10.9%)</td>
</tr>
<tr>
<td>RH</td>
<td>175 (81.8%)</td>
<td>69 (75%)</td>
</tr>
<tr>
<td>Mixed</td>
<td>29 (13.6%)</td>
<td>13 (14.1%)</td>
</tr>
<tr>
<td><strong>Injury Mechanism</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fall/hit by object</td>
<td>135 (63.1%)</td>
<td>48 (52.2%)</td>
</tr>
<tr>
<td>Fight/shaken</td>
<td>2 (0.9%)</td>
<td>3 (3.3%)</td>
</tr>
<tr>
<td>Multiple</td>
<td>10 (4.7%)</td>
<td>5 (5.4%)</td>
</tr>
<tr>
<td>Motor vehicle collision</td>
<td>14 (6.5%)</td>
<td>3 (3.3%)</td>
</tr>
<tr>
<td>Unknown</td>
<td>53 (24.8%)</td>
<td>33 (35.9%)</td>
</tr>
<tr>
<td><strong>Time Since Injury</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>3.22 (2.79)</td>
<td>3.23 (2.60)</td>
</tr>
</tbody>
</table>

Note: This table presents demographic and injury data for the Discovery set (n=214) and Replication set (n=92). The data includes mean and standard deviation for age, median age, sex distribution, pubertal stages, race/ethnicity, combined family income, handedness, injury mechanism, and time since injury. The table shows a comparison between the two sets, highlighting the variability in each category.
<table>
<thead>
<tr>
<th>Total TBIs</th>
<th>Unknown</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>53 (24.8%)</td>
<td>151 (70.6%)</td>
<td>9 (4.2%)</td>
<td>1 (0.5%)</td>
</tr>
<tr>
<td></td>
<td>33 (35.9%)</td>
<td>54 (58.7%)</td>
<td>5 (5.4%)</td>
<td>0 (0%)</td>
</tr>
</tbody>
</table>

Note: Participants with “Unknown” Injury Mechanism and Total TBIs reported sustaining a TBI but no mechanism of injury was endorsed.
ABCD Dataset (n=11,878)

Participants with history of mTBI (n=434) → No imaging data information (n=3)

Participants with imaging info (n=431) → Missing image information for all required images (n=1)

Participants with image information for minimal image requirements (n=430) → At least one required image missing (n=58)

Participants with DWI and/or B0 and/or T1 scans (n=372) → Missing or failed QC in at least one required image (n=13)

Participants with DWI and/or B0 and/or T1 scans that passed/had available QC (n=359) → Participants whose dicom -> nifti conversion failed (n=8)

Participants with complete data and nifti files (n=351) → Participants with mislabeled B0 file (n=6)

Final mTBI participant list preprocessing (n=345) → Participants who failed QC after processing (n=39)

Final mTBI participant list (n=306)