Supplemental Materials

Methods

Sample Characteristics

Fifty patients with pediatric mild traumatic brain injury (pmTBI; 12 – 18 years old) and 53 statistically matched (age and sex) healthy controls (HC) were enrolled into the study. Participants were removed from analyses for issues stemming from either visit. Two pmTBI failed screenings between assessment periods (pregnancy and repeat concussion). An additional 15 participants declined to perform the task (pmTBI = 5; HC = 3) or were removed from the scanner prior to performing the task (pmTBI = 5; HC = 2), which was the final one in the scanning sequence. Of the remaining participants, seven (pmTBI = 3; HC = 4) were excluded due to data acquisition errors. Ten participants (pmTBI = 3; HC = 7) were excluded for poor data quality following a visual inspection of end-tidal CO_2 ($ETCO_2$) and global grey matter signal for phases corresponding to room air and medical gas segments of the task. Four participants (pmTBI = 2; HC = 2) had inadequate overall fit ($r < .7$) between $ETCO_2$ and grey matter signal. No participants were identified as motion outliers (greater than 3 times interquartile range) for mean framewise displacement (FD) relative to their cohort. Thus, there were a total of 30 pmTBI (20 males; mean age 15.57±2.18) and 35 HC (21 males; 15.54±1.94) included during the final hypercapnia task analyses.

Finally, an additional 3 pmTBI patients and 4 HC were excluded as motion outliers during cerebral blood flow (CBF) analyses, leaving 27 pmTBI (18 males; mean age = 15.39±2.20) and 31 HC (17 males; mean age = 15.72±1.86) in this cohort.

Clinical and Behavioral Measures

A battery of clinical and neuropsychological measures was administered to pmTBI and HC at each visit. Measures included previous medical history, the Alcohol, Smoking and Substance Involvement Screening Test (ASSIST; WHO Group, 2002), self and parent reports of concussion symptom severity for both retrospective and current periods (Post-Concussion Symptom Inventory [PCSI]; Gioia et al., 2008; Gioia et al., 2009), Patient Reported Outcomes Measurement Information System (PROMIS) for sleep (Buysse et al., 2010), anxiety, and depression (Pilkonis et al., 2011), a brief pain rating (0-10 Likert scale; Farrar et al., 2001), self-report of Tanner stage of development (Kriz et al., 2016), Headache Impact Test (HIT-6; Kosinski et al., 2003), the Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997), Conflict and Behavioral Questionnaire (CBQ; Prinz et al., 1979), quality of life (Pediatric Quality of Life Inventory-PedsQL Generic Core; Varni et al., 1999), and the Glasgow Outcome Scale Extended (GOS-E; Beers et al., 2012) Pediatric Revision. Parental distress was measured with the Brief Symptom Inventory (BSI-18; Derogatis and Fitzpatrick, 2004).

The cognitive battery included tests of premorbid cognitive ability (Wide Range Achievement Test [WRAT-4]; Wilkinson and Robertson, 2006), a shortened measure of effort (Test of Memory Malingering – TOMMe10; Denning, 2012), the Cogstate battery (Cromer et al., 2015), and selected tests from the Delis-Kaplan Executive Function System (DKEFS; Delis et al., 2001) and Wechsler Intelligence Scales depending on initial age at assessment. Specifically, the Wechsler Adult Intelligence Scale – IV (WAIS-IV; Wechsler, 2008) was used for participants 16-18 year old at enrollment whereas the Wechsler Intelligence Scale for Children – V (WISC-V; Wechsler, 2014) was used for participants 12-15 year olds at enrollment. Composite measures of attention (DKEFS color-word interference conditions 1-3), processing speed (WAIS-IV/WISC-V digit symbol coding and symbol search), working memory (WISC-

V/WAIS-IV digit span backwards trial) and executive function (DKEFS trail making test condition 4, verbal fluency, color-word interference condition 4) were compiled to create specific cognitive domains. Cogstate battery reaction time and accuracy data underwent log and arcsine square root transformations, respectively, to generate computerized domain scores. Unadjusted central tendency measures for all clinical and neuropsychological measures (dependent on distribution) are presented in all tables.

Task Description

All participants completed a hypercapnia challenge modified from Lu et al. (2014) while undergoing BOLD-based imaging (Figure 1A). Briefly, participants were instructed to breath normally through a snorkel-like tube connected to a one-way non-rebreathing valve. Staff in the MR scanner environment were cued via in-house designed headphones to toggle a gas intake valve between room air (30±5 s) and medical grade gas mixture (5% CO_2, 21% O_2, BAL NO_2; 35 s) contained in a 50-liter Douglas bag. After an initial 15 s of room air, participants breathed in four fixed-length blocks (35 s) of gas mixture with a pseudorandom inter-block interval of room air (30±5 s) to allow better modelling of hemodynamic change. Gas mixture blocks are shorter than is typically used in BOLD CVR, but the efficacy of a truncated design is supported by previous literature (Bright et al., 2009) and our own initial pilot work. A 25ft sampling tube connected to a BIOPAC MP150 system and CO2100C module (BIOPAC Systems, Inc., Goleta, CA) outside the MR environment gathered continuous data on measured exhaled CO_2 (1000 Hz sampling frequency), which was monitored by the attending staff on a rear projected screen for real time quality assessment. The MP150 also logged image acquisition times to allow for temporal alignment between fMRI and CO_2 data. Importantly, participants had no indication of which air source they were receiving at any given time.

MR Imaging Parameters

All participants underwent imaging on a 3T Siemens Trio scanner with a 32-channel head coil. Foam padding was used to minimize head motion. A high resolution 5-echo Magnetization Prepared Rapid Acquisition Gradient Echo (MPRAGE) T_1–weighted [repetition time (TR) = 2530 ms; echo times (TE) = 1.64, 3.5, 5.36, 7.22, 9.08 ms; inversion time (TI) = 1200 ms; flip angle = 7°; number of excitations (NEX) = 1; slice thickness = 1 mm; field of view (FOV) = 256 mm; matrix size = 256×256; isotropic voxels = 1 mm] was collected in addition to a T_2–weighted sequence [TR = 15500 ms; TE = 77 ms; flip angle = 155°; NEX = 1; slice thickness = 1.5mm; FOV = 220 mm; matrix size = 192×192; voxel size = 1.15×1.1.5×1.5 mm]. Susceptibility weighted images (SWI) were collected using a single T_2-weighted gradient echo sequence [TR = 28 ms; TE = 20.0 ms; flip angle = 15; NEX = 1; slice thickness = 1.5 mm; FOV = 192×256; matrix size = 192×256; 88 interleaved slices; 1.00×1.00×1.50 mm voxels]. Fluid attenuated inversion recovery (FLAIR) data were collected using the following parameters [TR = 10380 ms; TE = 88.0 ms; TI = 2500 ms; flip angle = 140; NEX = 1; slice thickness = 3 mm; FOV = 256; matrix size = 320×320; 50 interleaved slices; 0.80×0.80×3.00 mm voxels]. These sequences were all reviewed by a board-certified neuroradiologist and all trauma-related findings are reported in Supplemental Table 3.

Functional data were acquired with a single-shot, gradient-echo echoplanar pulse sequence [TR = 460 ms; TE = 29 ms; flip angle = 44°; multiband acceleration factor = 8; NEX = 1; slice thickness = 3 mm; FOV = 248 mm; matrix size = 82×82] over a single run of the task, with fifty-six interleaved 3 mm slices acquired for whole-brain coverage (voxel size: 3.02×3.02×3.00 mm). A single run of resting state data was also collected. A reference image with multiband acceleration factor set to one was also acquired for each run to facilitate

registration with native T_1-weighted anatomical image due to improved grey-white matter contrast. To account for susceptibility artifacts in the gradient echo data, two spin-echo field mapping sequences [TR = 7220 ms; TE = 73 ms; flip angle = 90°; refocus flip angle = 180°; slice thickness = 3 mm; FOV = 248 mm; matrix size = 82×82; 56 interleaved slices; 3.02×3.02×3.00 mm voxels] with reversed phase encoding directions (A → P; P → A) were collected.

Finally, a pseudo-Continuous Arterial Spin Labeling (pCASL; 45 tagged/untagged images) sequence was acquired [TR = 4250 ms; TE = 11 ms; label offset = 90 mm; NEX = 1; slice thickness = 5 mm with 20% gap; bandwidth = 2790 Hz/Px; labeling duration = 1665 ms] with 20 interleaved slices for whole brain coverage (voxel size = 3.44×3.44×6.00 mm]. A proton density sequence was also acquired to estimate T_1 magnetization and scale CBF on a voxel-wise basis, with the post-labeling delay (PLD) and TR (5200 ms) being the only parameters that varied across the pCASL (PLD = 1800 ms) and PD scans (PLD = 3400 ms).

Physiological Analysis and Modelling

CO_2 data were initially processed in MATLAB (detailed in Figure 2; Mathworks, Inc., Natick, MA; version R2014) using a 125 ms rolling maximum filter. CO_2 data then underwent a post-collection recalibration based on solving a two point linear equation with observed room air baseline (ambient CO_2 levels) and observed medical gas block floor (5% CO_2 adjusted to altitude). The purpose of this was two-fold. Due to variations in acquisition templates between participants and visits, a series of proactive steps needed to be taken to ensure consistency in the data. For example, data in one version of the template were collected at percent CO_2 and were converted to mmHg CO_2 prior to processing. This two point solution also accounted for minor

drift in the signal collected by the CO2100C module between periodic maintenance and physical recalibration of the device.

The majority of peaks in data (i.e., end-tidal CO_2 [$ETCO_2$ peaks]) were identified algorithmically using the PeakFinder function (Yoder, 2009) hosted on the MATLAB File Exchange. Breath peaks were required to: 1) be separated by at least a 2 mmHg drop in CO_2; 2) exceed 15 mmHg CO_2; and 3) not be extreme relative to other peaks in $ETCO_2$ (i.e., within two times the interquartile range to filter aberrant data spikes). In the 15.7% of cases where peak modelling failed using standard parameters, data were visually inspected for compliance and parameters adjusted (e.g., smoothing or despiking window expanded).

After breath peaks were identified, interpolation between peaks was performed using a smoothing spline to create a smooth $ETCO_2$ vector. Data along this vector were then smoothed with a median filter (10 s window) and resampled to match image acquisitions (Figure 1B). This $ETCO_2$ regressor was fit against global grey matter signal (generated using subject-specific segmentation of T_1–weighted image; 60% probability of voxel being grey matter) using a family of offsets (7.82 to 18.4 seconds post-image collection; 1 to 24 image lag) to determine the optimal offset for each participant (Yezhuvath et al., 2009). This accounted for individual differences in physiology (gas exchange in lungs, blood to brain time, etc.) and the transit time of exhaled breath in the CO_2 sampling tube (a 7.36 second constant; Figure 1C). Participants with a correlation of less than 0.7 for the global fit (2 pmTBI; 2 HC) were excluded from further analyses.

Additionally, respiratory rates for the room air and medical gas experimental segments were determined algorithmically using average peak-to-peak times from the CO_2 data. Specifically, arrays were built which logged the time relative to the start of the scan at which

inspiration ceased and expiration occurred (for briefness, these will also be referred to as peaks). Peak times were allotted to either room air or medical gas epochs based on whether peak CO_2 values at that time were above or below the mean of all CO_2 peaks in the array. To avoid bias resulting from a straight binary classification, peaks falling in a between-epoch transitionary period within .25 times the interquartile range of the collective mean of CO_2 peaks were excluded from both epochs. Peak-to-peak times between consecutive breathes within the same epoch were calculated and then averaged to create a measure of respiratory rate. This calculation was not performed on data in which manual scalar adjustments were made due to concerns of peaks in medical gas epoch being misidentified as within the transitionary period.

MR Processing and Analysis

Functional imaging maps (AFNI's 3dfim+) were calculated using the family of time lagged $ETCO_2$ regressors to select both for the maximal fit and associated latency to maximal fit (Golestani et al., 2016; Liu et al., 2017). Voxel-wise CVR amplitude maps were also computed by dividing the beta coefficient of the maximum $ETCO_2$ regressor (β_{ETCO2}) by the model intercept (β_0) plus β_{ETCO2} times the minimum $ETCO_2$ value (Lu et al., 2014). This effectively results in a measure of change in BOLD signal per unit change in $ETCO_2$. Maps of the Z-transformed Pearson correlation coefficient at the index of maximal fit were generated in keeping with latency maps dependent on this correlation and to increase sensitivity by normalizing the data (Svaldi et al., 2017), though directionality of between-group effects should not change in fit maps relative to amplitude maps. Because of the extended block design of the experiment, motion parameters were not included in the model (Mayer et al., 2019).

Results

Clinical and Neuropsychological Analyses

Analysis of retrospective data showed significantly increased sleep issues ($t_{63} = 2.36$; $p = 0.022$) and child ratings of behavior ($Z_U = 2.01$; $p = 0.044$) in pmTBI relative to HC, but groups were statistically similar otherwise.

For secondary clinical measures, generalized estimating equations [Group (pmTBI vs. HC) × Time (Sub-Acute vs. Early Chronic)] were performed with retrospective ratings as covariate when collected. A Group×Time interaction was observed for anxiety (*Wald*-$\chi^2 = 7.05$; $p = 0.008$), headache (*Wald*-$\chi^2 = 10.61$; $p = 0.001$) and pain (*Wald*-$\chi^2 = 14.26$; $p < 0.001$), with significant differences observed between groups at the sub-acute (pmTBI > HC; all *p's* < 0.005) but not early chronic phases. There was a main effect of Group for sleep disturbances (*Wald*-$\chi^2 = 4.25$; $p = 0.039$), with pmTBI reporting increased symptoms relative to HC. Depression did not differ between groups (*p*'s > 0.05).

For primary cognitive measures, main effects of Time (early chronic > sub-acute) were observed for traditional (i.e., paper and pencil) domains of attention (*Wald*-$\chi^2 = 5.63$; $p = 0.018$) and processing speed (*Wald*-$\chi^2 = 23.91$; $p < 0.001$), indicating better performance during early chronic testing. There were no main effects of Time for reaction time or accuracy data on computerized tests, nor were interactions on accuracy data significant (all *p's* > 0.10). However, processing speed (*Wald*-$\chi^2 = 3.99$; $p = 0.046$), but not attention, showed a significant main effect of Group, with pmTBI exhibiting lower accuracy relative to HC.

There were no significant interactions or main effects (*p*'s > 0.05) on secondary measures of effort (TOMMe10) or premorbid reading ability (WRAT-4). A significant main effect of Time was observed for the secondary measure of executive functioning (*Wald*-$\chi^2 = 36.86$; $p < 0.001$), again indicating better performance during early chronic testing, while working memory showed

no main effects or interaction (p's > 0.05). For reaction time on secondary computerized measures, a main effect of Time (sub-acute > early chronic) was observed for both working memory (*Wald*-χ^2 = 6.39; p = 0.011) and learning (*Wald*-χ^2 = 3.93; p = 0.048). Accuracy effects were null across both secondary computerized measures (p's > 0.05).

Functional Imaging –Main Effects of Time

A main effect of Time on CVR amplitude data was observed. In all significant areas, there was a decrease in the magnitude of response to CVR at the early chronic relative to the sub-acute visit. As opposed to the main effects of Group, this change was limited almost exclusively to grey matter (cortical, sub-cortical, and cerebellar), with much of the areas showing difference in the bilateral auditory and sensorimotor cortices (Supplemental Figure 3). Leung and colleagues (2016) previously described increasing CVR amplitude up to 14.7 years old followed by decreasing amplitude. However, follow-up analyses indicated that age was not associated in either a linear or polynomial fashion with CVR amplitude at the initial visit in either HC or participants as a whole, nor was change in age between visits associated with change in CVR amplitude between visits (all p's > 0.10). Time effects for maximal fit and latency to maximal fit were null.

Relationship between Cerebrovascular Deficits and Clinical Outcomes

A binary logistic regression determined whether differences in cerebrovascular fit or delay (weighted means of all significant areas) classify pmTBI from HC. Results indicated that both weighted fit (β = 45.77; *Wald*-χ^2 = 8.01; p = 0.005; OR not reported because of scale) and delay (β = 3.19; *Wald*-χ^2 = 9.59; p = 0.002; OR = 24.40) were significant predictors in classifying patients (90.0%; 27/30) from HC (91.4%; 32/35) with a total accuracy of 90.8% (Δ =

37.0% relative to baseline model). Amplitude ROI proved less effective when used in place of fit at predicting membership of the pmTBI group (80.0%; 24/30), with inclusion of all three weighted ROI means in the model classifying only one additional HC correctly (total accuracy = 92.3%) relative to the latency/fit model.

References

Beers, S.R., Wisniewski, S.R., Garcia-Filion, P., Tian, Y., Hahner, T., Berger, R.P., Bell, M.J., and Adelson, P.D. (2012). Validity of a pediatric version of the Glasgow Outcome Scale-Extended. J Neurotrauma *29*, 1126-1139.

Bright, M.G., Bulte, D.P., Jezzard, P., and Duyn, J.H. (2009). Characterization of regional heterogeneity in cerebrovascular reactivity dynamics using novel hypocapnia task and BOLD fMRI. Neuroimage *48*, 166-175.

Buysse, D.J., Yu, L., Moul, D.E., Germain, A., Stover, A., Dodds, N.E., Johnston, K.L., Shablesky-Cade, M.A., and Pilkonis, P.A. (2010). Development and validation of patient-reported outcome measures for sleep disturbance and sleep-related impairments. Sleep *33*, 781-792.

Cromer, J.A., Schembri, A.J., Harel, B.T., and Maruff, P. (2015). The nature and rate of cognitive maturation from late childhood to adulthood. Front Psychology *6*, 704.

Delis,D.C., Kaplan,E., and Kramer,J.H. (2001). Delis-Kaplan executive function system (D-KEFS). Psychological Corporation.

Denning, J.H. (2012). The efficiency and accuracy of the Test of Memory Malingering trial 1, errors on the first 10 items of the test of memory malingering, and five embedded measures in predicting invalid test performance. Arch Clin Neuropsychol *27*, 417-432.

Derogatis, L.R., and Fitzpatrick, M. (2004). The SCL-90-R, the Brief Symptom Inventory (BSI), and the BSI-18.

Farrar, J.T., Young, J.P., Jr., LaMoreaux, L., Werth, J.L., and Poole, R.M. (2001). Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. Pain *94*, 149-158.

Gioia, G.A., Collins, M., and Isquith, P.K. (2008). Improving identification and diagnosis of mild traumatic brain injury with evidence: psychometric support for the acute concussion evaluation. J Head Trauma Rehabil *23*, 230-242.

Gioia, G.A., Schneider, J.C., Vaughan, C.G., and Isquith, P.K. (2009). Which symptom assessments and approaches are uniquely appropriate for paediatric concussion? Br J Sports Med *43 Suppl 1*, i13-i22.

Golestani, A.M., Wei, L.L., and Chen, J.J. (2016). Quantitative mapping of cerebrovascular reactivity using resting-state BOLD fMRI: Validation in healthy adults. Neuroimage *138*, 147-163.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: a research note. J Child Psychol Psychiatry *38*, 581-586.

Kosinski, M., Bayliss, M.S., Bjorner, J.B., Ware, J.E., Jr., Garber, W.H., Batenhorst, A., Cady, R., Dahlof, C.G., Dowson, A., and Tepper, S. (2003). A six-item short-form survey for measuring headache impact: the HIT-6. Qual Life Res *12*, 963-974.

Kriz, P.K., Stein, C., Kent, J., Ruggieri, D., Dolan, E., O'Brien, M., and Meehan, W.P., III (2016). Physical Maturity and Concussion Symptom Duration among Adolescent Ice Hockey Players. J Pediatr *171*, 234-239.

Leung, J., Kosinski, P.D., Croal, P.L., and Kassner, A. (2016). Developmental trajectories of cerebrovascular reactivity in healthy children and young adults assessed with magnetic resonance imaging. J Physiol *594*, 2681-2689.

Liu, P., Li, Y., Pinho, M., Park, D.C., Welch, B.G., and Lu, H. (2017). Cerebrovascular reactivity mapping without gas challenges. Neuroimage *146*, 320-326.

Lu, H., Liu, P., Yezhuvath, U., Cheng, Y., Marshall, O., and Ge, Y. (2014). MRI mapping of cerebrovascular reactivity via gas inhalation challenges. J Vis Exp *94*, e52306.

Mayer, A.R., Ling, J.M., Dodd, A.B., Shaff, N.A., Wertz, C.J., and Hanlon, F.M. (2019). A comparison of denoising pipelines in high temporal resolution task-based functional magnetic resonance imaging data. Hum Brain Mapp *40*, 3843-3859.

Pilkonis, P.A., Choi, S.W., Reise, S.P., Stover, A.M., Riley, W.T., and Cella, D. (2011). Item banks for measuring emotional distress from the Patient-Reported Outcomes Measurement Information System (PROMIS(R)): depression, anxiety, and anger. Assessment *18*, 263-283.

Prinz, R.J., Foster, S., Kent, R.N., and O'Leary, K.D. (1979). Multivariate assessment of conflict in distressed and nondistressed mother-adolescent dyads. J Appl Behav Anal *12*, 691-700.

Svaldi, D.O., McCuen, E.C., Joshi, C., Robinson, M.E., Nho, Y., Hannemann, R., Nauman, E.A., Leverenz, L.J., and Talavage, T.M. (2017). Cerebrovascular reactivity changes in asymptomatic female athletes attributable to high school soccer participation. Brain Imaging Behav *11*, 98-112.

Varni, J.W., Seid, M., and Rode, C.A. (1999). The PedsQL: measurement model for the pediatric quality of life inventory. Med Care *37*, 126-139.

Wechsler, D. (2014). Wechsler intelligence scale for children – Fifth edition (WISC-V): Technical and interpretive manual. Bloomington, MN: Pearson Clinical Assessment.

Wechsler,D. (2008). Wechsler adult intelligence scale-fourth. San Antonio: Pearson.

WHO Group (2002). The alcohol, smoking and substance involvement screening test (ASSIST): development, reliability and feasibility. Addiction *97*, 1183-1194.

Wilkinson,G.S. and Robertson,G.J. (2006). WRAT 4: Wide range achievement test; professional manual. Psychological Assessment Resources, Incorporated.

Yezhuvath, U.S., Lewis-Amezcua, K., Varghese, R., Xiao, G., and Lu, H. (2009). On the assessment of cerebrovascular reactivity using hypercapnia BOLD MRI. NMR Biomed *22*, 779-786.

Yoder, Nathanael. (11-6-2009). PeakFinder (http://www.mathworks.com/matlabcentral/fileexchange/25500). Retrieved August 22, 2016. MATLAB Central File Exchange.

Supplemental Table 1: Sample size and demographic information for various imaging modalities.

Exclusion Reason	pmTBI N (sex, age)	HC N (sex, age)
Enrolled	50 (30 M, 15.44±2.14)	53 (34 M, 15.22±2.00)
Failed EC screening	48 (29 M, 15.44±2.15)	53 (34 M, 15.22±2.00)
Didn't perform task	38 (24 M, 15.41±2.18)	48 (30 M, 15.44±1.93)
Acquisition errors	35 (23 M, 15.45±2.18)	44 (27 M, 15.42±1.94)
Visual quality exclusion	32 (20 M, 15.41±2.22)	37 (23 M, 15.47±1.91)
GM fit exclusion	30 (20 M, 15.57±2.18)	35 (21 M, 15.54±1.94)
Task: motion outlier	30 (20 M, 15.57±2.18)	35 (21 M, 15.54±1.94)
CBF: motion outlier	27 (18 M, 15.39±2.20)	31 (17 M, 15.72±1.86)

Notes: Total N after exclusion (Number of males [M], mean age ± standard deviation). Exclusions are cumulative through Task motion outliers, with the cerebral blood flow (CBF) sample size derived from those who were included in task analysis. EC = early chronic; GM = grey matter; HC = healthy control; pmTBI = pediatric mild traumatic brain injury.

Supplemental Table 2: Retrospective measures.

	Outcome	HC	pmTBI
PCSI	P	4(1-10)	5(0-15.75)
PCSI (Parent)	P	1.5(0-6.75)	2(0-6)
PROMIS Sleep *	S	14.34±3.93	17.1±5.47
PROMIS Anxiety	S	3(1-6)	3(0-8)
PROMIS Depression	S	1(0-3.5)	2(0-5.5)
Pain Scale	S	0(0-1)	0(0-3)
HIT-6	S	45(40-52)	50(42-54)
CBQ *	P	0(0-1.5)	1(0-3)
CBQ (Parent)	P	1(0-2)	1(0-2)
SDQ (Parent)	S	4.5(2-8.75)	6(3.5-8)
PedsQL	P	89.13(80.43-92.39)	83.7(77.72-87.77)
PedsQL (Parent)	P	87.5(79.62-95.65)	85.87(70.11-91.85)

Notes: CBQ = Children's Behavior Questionnaire; HC = healthy control; HIT-6 = Headache Impact Test; P = primary; PCSI = Post-Concussion Symptom Inventory; PedsQL = Pediatric Quality of Life Inventory; pmTBI = pediatric mild traumatic brain injury; PROMIS = Patient-Reported Outcomes Measurement Information System; S = secondary; SDQ = Strengths and Difficulties Questionnaire. Data are either formatted at mean ± standard deviation or median (interquartile range). Asterisks indicate significance.

Supplemental Table 3: Summary of traumatic CT/MRI Findings

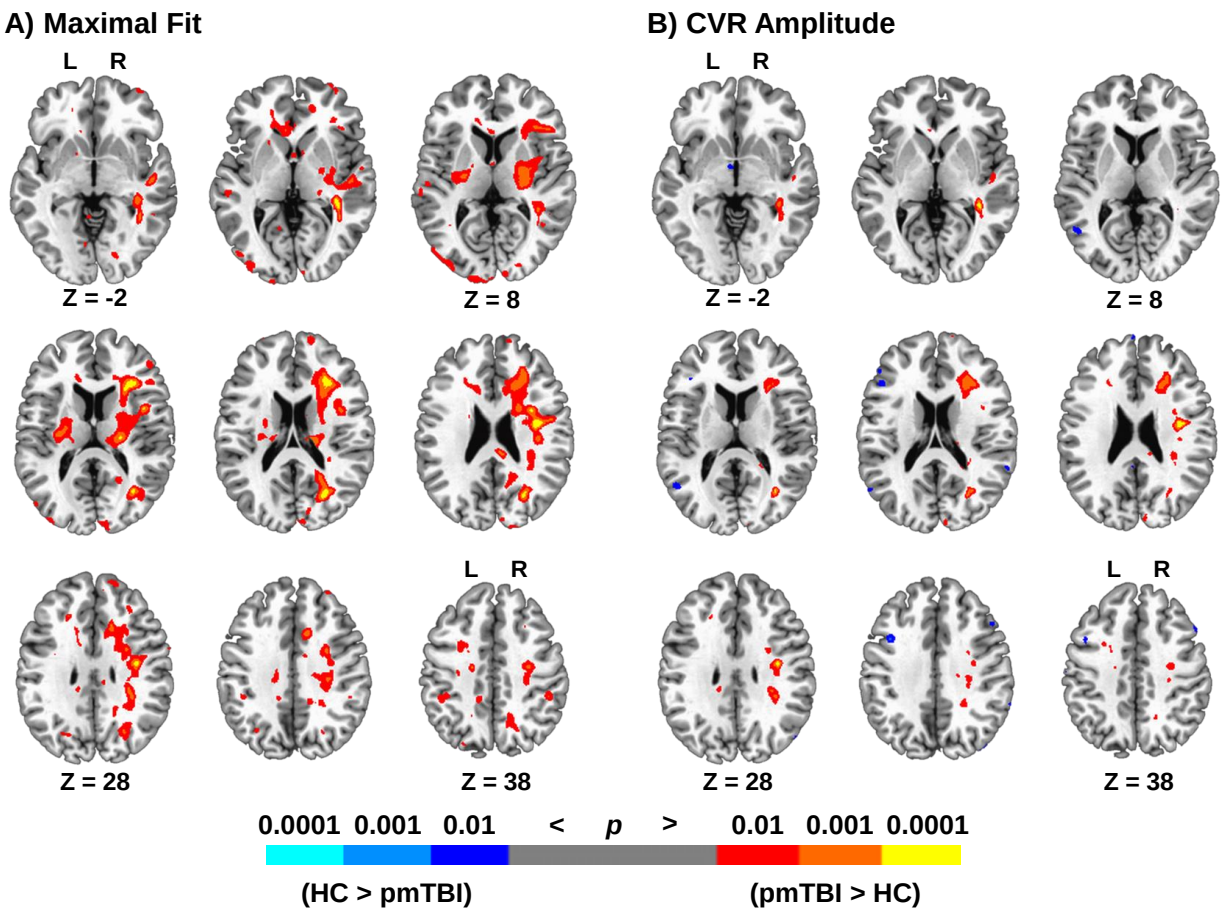
Age	Sex	Injury Mechanism	CT Finding	MRI Finding
13	M	Fall while skateboarding	CT notes unavailable, but referral indicated: 'CT […] showed left temporal skull fracture.'	Left mastoid temporal bone fracture; 1, left parietal, small (27x6 mm) subdural hematoma; high right frontal white matter contusion (6mm).
16	F	Motor vehicle accident	Extra-axial fluid collections over the lateral/temporal cerebral convexities bilaterally, measuring up to 3 mm on the right and 2 mm on the left. No midline shift. Basilar cisterns are patent. No hydrocephalus. Incidentally noted cavum septum et vergae. No large acute ischemic infarction is identified. Dural venous sinuses appear normal. Intraorbital contents are normal. Left frontal scalp, left temporal scalp, and left cheek soft tissue contusions. No acute osseous abnormality. Paranasal sinuses and mastoid air cells are clear.	2, left frontal and temporal subcortical white matter hemorrhagic traumatic axonal injury, 1-2 cm longitudinal.
14	M	Fall during bull riding	Intraparenchymal hemorrhage is noted in the white matter of the frontal lobe superior frontal gyrus near the vertex with surrounding mild edema is measured 1.7 x 0.7 x 2.0 cm (AP by transaxial by cc). Trace hyperdensity favored to represent subarachnoid hemorrhage is noted near the vertex on image #142/169. Punctate hemorrhage is noted in the right anterior insula. No hydrocephalus or herniation.	2, anterior right frontal subcortical and high right insular cortex contusions (20x10mm and 8x6mm); 5, left insular cortex and right medial temporal cortex and along left fornix, and bilateral midbrain, punctate lesions mostly oriented in AP linear fashion; Some of the lesions described as DAI could be grey matter micro-bleeds.

Notes: F = female; M = male; NA = not applicable.

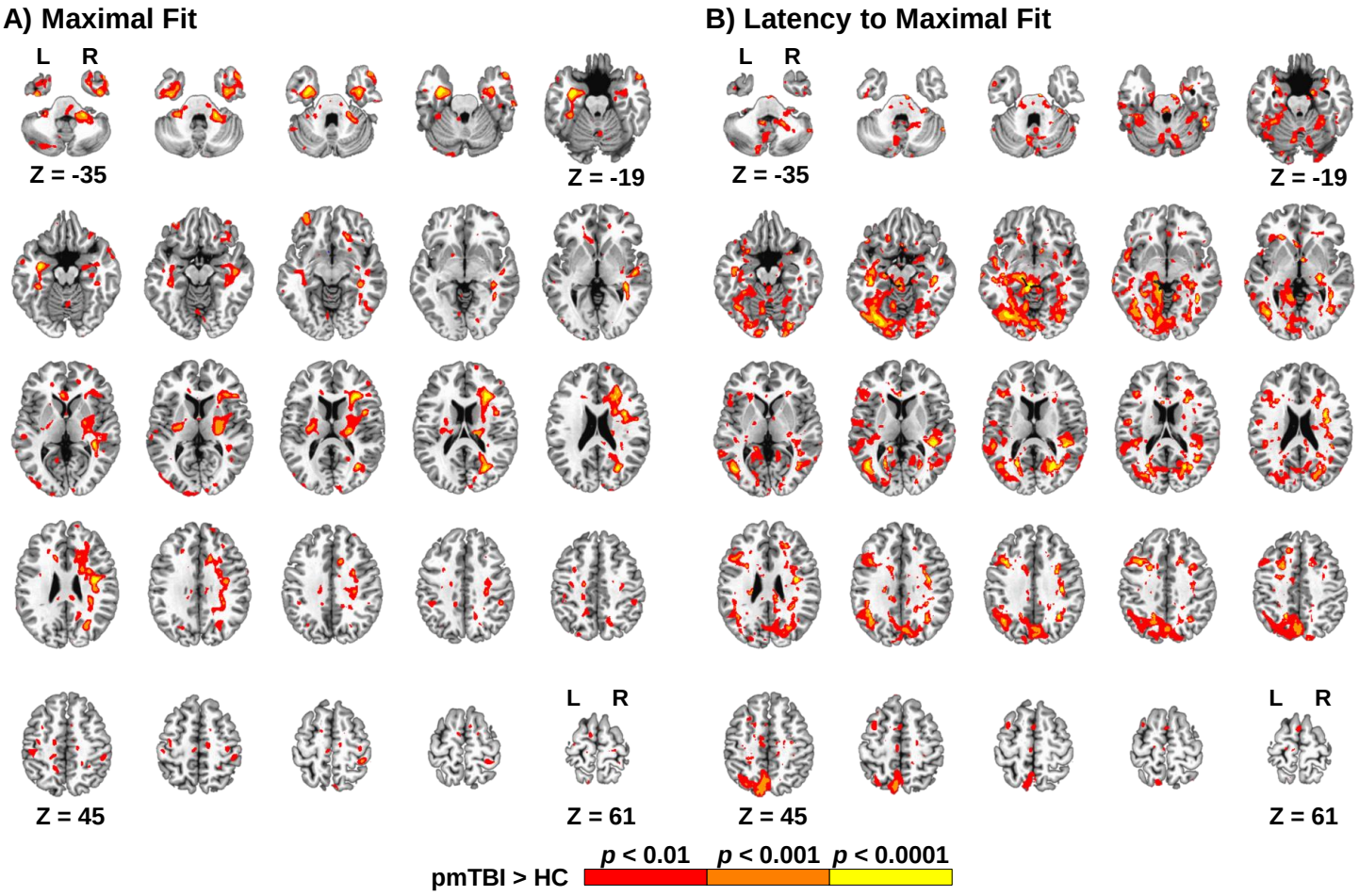
Supplemental Table 4: Individual demographic and self-reported injury characteristics of pmTBI sample

Sex	Age	DPI	Injury Location		Mechanism of Injury	SRR Injury	SRR Category	Activity
			L/M/R	A/MC/P				
F	12	9	R	A	Struck by object	Yes	Contact	Soccer
F	12	6	R	A	Struck by person	Yes	Contact	Basketball
F	12	6	R	P	MVC	Yes	Contact	Extreme Sports
F	14	9	M	P	Fall	Yes	Contact	Rugby
F	16	10	M	P	MVC	No	--	--
F	16	7	--	--	MVC	No	--	--
F	16	9	L	A	MVC	No	--	--
F	17	7	L	A	MVC	No	--	--
F	18	7	R	A	MVC	No	--	--
F	18	7	R	A	MVC	No	--	--
M	12	7	L	A	Fall	Yes	Contact	Soccer
M	12	4	L	P	Struck by object	Yes	Contact	Soccer
M	13	3	R	A	Bicycle-related	Yes	Limited contact	Bicycling
M	13	9	R	A	Struck by person	Yes	Contact	Football
M	13	7	L	P	Fall	Yes	Limited contact	Skateboarding
M	13	8	M	P	Bicycle-related	Yes	Limited contact	Bicycling
M	14	6	R	A	Fall	Yes	Contact	Rodeo
M	14	7	L	MC	Fall	Yes	Limited contact	Skateboarding
M	15	9	M	MC	Assault	No	--	--
M	15	5	M	P	Struck by person	Yes	Contact	Soccer
M	15	9	M	P	Fall	Yes	Contact	Rodeo
M	16	4	R	A	Struck by person	Yes	Contact	Football
M	16	7	L	MC	Fall	No	--	--
M	16	8	M	P	Fall	Yes	Limited contact	Skateboarding
M	16	3	M	A	MVC	No	--	--
M	17	9	L	P	MVC	No	--	--
M	18	1	M	P	Struck by object	No	--	--
M	18	7	L	P	Struck by person	Yes	Contact	Basketball
M	18	8	R	P	MVC	No	--	--
M	18	9	L	A	Bicycle-related	Yes	Limited contact	Bicycling

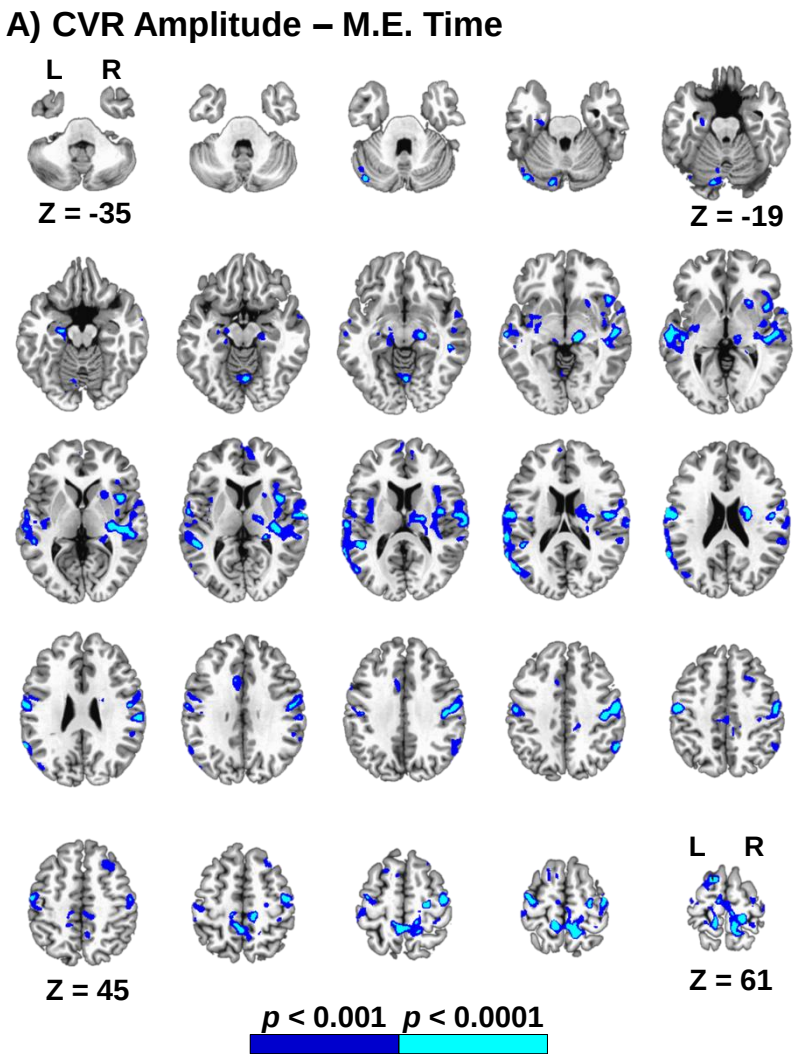
Notes: F = female; M = male; DPI = days post injury; L/M/R = left/medial/right; A/MC/P = anterior/mid-coronal/posterior; MVC = motor vehicle collision; SRR = sports or recreation related.



Supplemental Figure 1: This figure depicts maps of main effect of Group for maximal fit (Panel A) and amplitude of cerebrovascular reactivity (CVR; Panel B) during a hypercapnia task following pediatric mild traumatic brain injury (pmTBI) and in healthy controls (HC). Importantly, this data is displayed with no volume threshold and only a small statistical threshold. Regions with increased fit or amplitude in pmTBI relative to HC appear in warm colors, whereas regions with decreased fit or amplitude (HC > pmTBI) appear in cool colors. Select axial (Z) slices are displayed at 5 mm intervals according to the Talairach atlas with the right (R) and left (L) hemispheres denoted.



<u>Supplemental Figure 2</u>: This figure displays brain areas showing a significant main effect of Group (pediatric mild traumatic brain injury [pmTBI] > healthy controls [HC]) for increased maximal voxel-wise fit (Panel A) and latency to maximal fit (Panel B) to the $ETCO_2$ regressor matrix (red: $p < 0.01$; orange: $p < 0.001$; yellow: $p < 0.0001$). Regions are purposely illustrated with no minimum volume threshold and a lower statistical threshold to demonstrate the spatial independence of the regions in each analysis. Select axial (Z) slices are displayed at 4 mm intervals according to the Talairach atlas with the right (R) and left (L) hemispheres denoted.



<u>Supplemental Figure 3</u>: Panel A displays brain areas showing a significant main effect (M.E.) of Time for amplitude of cerebrovascular reactivity (CVR) during a hypercapnia task following pediatric mild traumatic brain injury. Regions exhibiting decreased magnitude of response to hypercapnia from the sub-acute to the early chronic visit are represented in cool colors (blue: $p < 0.001$ cyan: $p < 0.0001$) following correction for false positives. Select axial (Z) slices are displayed at 4 mm intervals according to the Talairach atlas with the right (R) and left (L) hemispheres denoted.