Comparison of Methods for Classifying Persistent Post-Concussive Symptoms in Children

Andrew R. Mayer,[1–4] David D. Stephenson,[1] Andrew B. Dodd,[1] Cidney R. Robertson-Benta,[1]
Sharvani Pabbathi Reddy,[1] Nicholas A. Shaff,[1] Keith Owen Yeates,[5–7] Harm J. van der Horn,[1,8]
Christopher J. Wertz,[1] Grace Park,[9] Scott J. Oglesbee,[9] Edward J. Bedrick,[10]
Richard A. Campbell,[2] John P. Phillips,[1,4] and Davin K. Quinn[2]

Abstract

Pediatric mild traumatic brain injury (pmTBI) has received increased public scrutiny over the past decade, especially regarding children who experience persistent post-concussive symptoms (PPCS). However, several methods for defining PPCS exist in clinical and scientific literature, and even healthy children frequently exhibit non-specific, concussive-like symptoms. Inter-method agreement (six PPCS methods), observed misclassification rates, and other psychometric properties were examined in large cohorts of consecutively recruited adolescent patients with pmTBI ($n = 162$) 1 week and 4 months post-injury and in age/sex-matched healthy controls (HC; $n = 117$) at equivalent time intervals. Six published PPCS methods were stratified into Simple Change (e.g., *International Statistical Classification of Diseases and Related Health Problems*, 10th revision [ICD-10]) and Standardized Change (e.g., reliable change indices) algorithms.

Among HC, test-retest reliability was fair to good across the 4-month assessment window, with evidence of bias (i.e., higher symptom ratings) during retrospective relative to other assessments. Misclassification rates among HC were higher (>30%) for Simple Change algorithms, with poor inter-rater reliability of symptom burden across HC and their parents. A 49% spread existed in terms of the proportion of pmTBI patients "diagnosed" with PPCS at 4 months, with superior inter-method agreement among standardized change algorithms. In conclusion, the self-reporting of symptom burden is only modestly reliable in typically developing adolescents over a 4-month period, with additional evidence for systematic bias in both adolescent and parental ratings. Significant variation existed for identifying pmTBI patients who had "recovered" (i.e., those who did not meet individual criteria for PPCS) from concussion across the six definitions, representing a considerable challenge for estimating the true incidence rate of PPCS in published literature. Although relatively straightforward to obtain, current findings question the utility of the most commonly used Simple Change scores for diagnosis of PPCS in clinical settings.

Keywords: pediatric mild traumatic brain injury; persistent post-concussive symptoms

Introduction

PUBLIC CONCERN IS GROWING regarding the increased incidence and potential long-term effects of pediatric mild traumatic brain injury (pmTBI), used synonymously here with concussion.[1–3] In 2013, children aged 14 years and younger accounted for 640,000 TBI-related emergency department (ED) visits in the United States, 70–90% of which were classified as pmTBI.[4,5] It has been suggested that approximately 30% of patients with pmTBI exhibit persistent post-concussive symptoms (PPCS) 1 month post-injury.[6] However, multiple methodologies exist for defining PPCS[7] and few studies have directly compared their psychometric properties[8] or longitudinally examined the rate of healthy adolescent controls (HC) who would be misdiagnosed with PPCS. The latter is critical given that approximately 20% of HC endorse symptoms consistent with PPCS, with higher proportions among individuals with learning disabilities, attention deficit hyperactive disorder, or mental health disorders.[9,10]

[1]The Mind Research Network/Lovelace Biomedical and Environmental Research Institute, Albuquerque, New Mexico, USA.
[2]Department of Psychiatry and Behavioral Sciences, [3]Department of Psychology, [4]Department of Neurology, and [9]Department of Emergency Medicine, University of New Mexico, Albuquerque, New Mexico, USA.
[5]Department of Psychology, [6]Alberta Children's Hospital Research Institute, and [7]Hotchkiss Brain Institute, University of Calgary, Calgary, Alberta, Canada.
[8]Department of Neurology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands.
[10]Department of Epidemiology and Biostatistics, University of Arizona, Tucson, Arizona, USA.

A large barrier for the field is the lack of a consensus definition on what symptoms (i.e., number and type) and what duration constitute PPCS.[7] Most clinical studies rely on Simple Change scores for determining PPCS. Specifically, previous studies have utilized three or more new symptoms at 2 weeks[8,11] or 1 month post-injury.[6,12] Others[13] have defined PPCS as a positive change of at least 2 points in a single symptom after 3 months, or utilized a variant of this method.[14] Critically, the specification of ''new'' or ''worsening'' symptoms inherently requires the availability of either baseline (typically unavailable) or retrospective, pre-injury ratings. Retrospective ratings are deemed to be more accurate if they are obtained closer to the actual injury date.[10] Adults with mTBI exhibit evidence of reporting bias compared with HC in the form of lower pre-injury symptoms.[15] More sophisticated statistical methods for measuring clinically meaningful symptom change based on standardization against control sample ratings, such as regression-based (RB)[16] or reliable change indices (RCI),[17] have also been used to determine PPCS (hereafter referred to as Standardized Change methods).[18–20]

Basic psychometric properties such as test-retest and inter-rater reliability are also critical for determining clinically meaningful symptom change.[21–24] The current study therefore examined agreement between six published Simple Change[6,13] or Standardized Change[16,17] methods (Supplementary Table S1) for classification of PPCS in pmTBI patients and HC. The a priori prediction was that the Standardized Change methods would minimize misclassification rates in HC and provide greater inter-method agreement. The reliability of symptom reporting at 1 week and 4 months post-injury was also examined. The focus of the test-retest reliability analyses were on HC given that symptoms were expected to change following pmTBI. Parental ratings of symptoms were also assessed given concerns about validity of self-report in pediatric samples,[21,24] and the utilization of parent-report rather than self-report ratings for defining PPCS.[6] Finally, between group (HC vs. pmTBI patients) comparisons of PPCS and other symptom scales were conducted to statistically characterize each group and determine whether there was significant change over time in the injured group.

Methods

Participants

Children with pmTBI (12–18 years of age) were consecutively recruited from local ED and urgent care settings between July 2016 and April 2019 for this prospective cohort study. Participants were subsequently evaluated in an outpatient setting at subacute (SA; approximately 1 week post-injury) and early chronic (EC; approximately 4 months) injury phases, corresponding to typical follow-up windows for patient care. Inclusion criteria were based on American Congress of Rehabilitation Medicine and Zurich Concussion in Sport Group[25] guidelines. Specifically, all pmTBI patients experienced head trauma resulting in Glasgow Coma Scale scores ≥13, an alteration in mental status or at least two new symptoms, loss of consciousness (if present) <30 min, and post-traumatic amnesia (if present) limited to 24 hours. HC were recruited from the local community through fliers and word of mouth, and were evaluated at equivalent time-points.

Exclusion criteria were the same for both pmTBI patients and HC: 1) a history of a) neurological diagnosis, b) previous moderate or severe TBI with >30 min loss of consciousness, c) developmental disorder [autism spectrum disorder or intellectual disability], d) any psychiatric disorders other than adjustment disorder, or e) substance abuse/dependence; or 2) a non-English-speaking child or guardian. HC were also excluded if diagnosed with attention deficit hyperactivity disorder or a learning disability. Urine-based

drug screens were conducted for all participants at both visits. Any positive result was exclusionary for HC. Children with pmTBI were not excluded if they tested positive for marijuana use ($n = 4$) due to concerns about the use of this substance for treatment of symptoms (self-medication or otherwise). With the exception of one measure (see Supplementary Appendix), overall group-wise results were fundamentally unchanged when marijuana users were removed from principal analyses. All participants provided informed consent or assent according to institutional guidelines at the University of New Mexico School of Medicine.

Procedures

Participants rated symptom severity retrospectively (i.e., for the month prior to injury) and at the SA and EC assessments using a modified version of the Post-Concussion Symptom Inventory (PCSI; see Supplementary Appendix S1). The total PCSI score was the primary outcome variable, whereas subscale (Physical, Cognitive, Emotional, and Fatigue) ratings were secondary. Additional secondary self-report symptom inventories included Patient Reported Outcomes Measurement Information System (PROMIS) scales for sleep, anxiety, and depression, and the Headache Impact Test (HIT-6). For all symptom inventories, retrospective ratings were collected only at the SA assessment in the current study to minimize reporting bias.[10] All participants also completed a semi-structured questionnaire about history of previous head injuries and a self-reported Tanner stage of development to ensure equivalent sexual development across groups.

PPCS definitions

A brief synopsis of the methods used to calculate symptom burden at both the SA and EC assessments is provided in the following section, with full mathematical details in Supplementary Table S1. Symptoms at the SA time-point were operationally defined as post-concussive symptoms (PCS), whereas symptoms still present at the EC phase were operationally defined as PPCS. Simple Change methods included the ICD-10 (i.e., at least 1-point increase in at least three symptoms)[6] and algorithms as reported by Smyth and colleagues (i.e., ≥2-point increase in at least one symptom),[13] and were based solely on change from retrospective ratings. Four Standardized Change methods utilized HC ratings as a reference in calculating both PCS and PPCS burden in conjunction with a standardized cutoff ($z = 1.64$). RCI was calculated based on the original Jacobson and Traux formula[17] as practice effects were not anticipated.[26]

For the second method, SA/EC were regressed separately on reported retrospective ratings within HC, with the derived coefficients used to determine significant deviations between observed and predicted SA/EC ratings for both groups.[16] Symptom burden was also estimated by summing PCSI ratings, calculating the common logarithm (Log_{10}) to reduce skew, standardizing (z-scores) based on HC ratings (classification cutoff, $z \approx 1.64$) and adjusting for distributional bias.[27] Finally, the difference between SA/EC and retrospective ratings were also summed, standardized (z-scores), and corrected for distributional bias. Based on cutoffs derived within each of the published methods, every individual from both the pmTBI and HC groups was then operationally classified as being either symptomatic (PCS or PPCS depending on time-point) or normal/''recovered.''

Statistical analysis

Within-subject analyses were performed with generalized estimating equations (GEE) using negative binomial or gamma distributions. Between group comparisons were performed with either generalized estimating equations and/or generalized linear models using the same distributions to better characterize recovery from

injury in the pmTBI group. Intraclass correlation coefficients (ICC)[28] were used to examine both test-retest reliability (ICC[2,1]) and inter-rater reliability (ICC[2,k]; self vs. parent). Gwet's AC1 estimation[29] evaluated agreement among non-independent PCS/PPCS classification methods. All reliability estimates were categorized as poor (≤ 0.39), fair (0.40–0.59), good (0.60–0.74), or excellent (≥ 0.75) based on previous guidelines.[30] Finally, confidence intervals were calculated using the Clopper-Pearson method[31] based on an assumed false-positive rate (i.e., misclassification; operationally defined as 5%) within HC.

Results

Participants

A total of 165 pmTBI patients and 120 statistically matched (age/sex) HC were included in the study. Six individuals in total were excluded (see Supplementary Appendix S1), resulting in a final sample of 162 pmTBI patients (77 females; age 14.9 ± 1.9 years; 7.4 ± 2.2 days post-injury) and 117 HC (55 females; age 15.0 ± 2.0 years). For the purposes of the current study, 141 of 162 pmTBI patients and 104 of 117 HC were eligible for the 4-month follow-up visit (i.e., a larger parent study is ongoing). Of these participants, 119 pmTBI patients (84.4% retention; 124.2 ± 12.9 days post-visit) and 94 HC (90.4% retention; 122.7 ± 11.3 days post-visit) completed their 4-month visit.

Group by time comparisons

Groups did not differ in terms of handedness, age, self-reported Tanner stage of development, biological sex, or previous head injuries (all p's > 0.10). See Table 1 for descriptive parameters of primary and secondary measures across all assessment periods and relevant sample sizes. Retrospective self-report (Wald-χ^2 = 30.08; $p < 0.001$) and parent-report (Wald-χ^2 = 14.48; $p < 0.001$) PCSI ratings were greater for pmTBI patients relative to HC, as well as all self-report retrospective ratings of secondary measures (all p's < 0.05).

A significant Group × Time interaction was observed for both self-report (Wald-χ^2 = 16.86; $p < 0.001$; Fig. 1A) and parent-report (Wald-χ^2 = 21.93; $p < 0.001$; Fig. 1B) PCSI ratings. Elevated symptom load existed at both SA and EC assessments (pmTBI patients > HC; all p's ≤ 0.001), but was of greater magnitude at SA (incidence risk ratio [IRR]: self = 4.21; parent = 6.89) relative to EC (IRR: self = 1.92; parent = 2.42). Secondary analyses indicated that sleep disturbance exhibited a similar interaction with a greater symptom magnitude (pmTBI patients > HC; Fig. 1D) at SA relative to EC (IRR = 1.43 vs. 1.26) assessment. In contrast, Group × Time interactions for anxiety (Fig. 1E) and headaches (Fig. 1C) indicated

significant group differences (pmTBI patients > HC) at SA only (p's ≤ 0.001; anxiety IRR = 2.52; headache IRR = 1.14). A main effect of Group was present for depression ($p < 0.001$; IRR = 1.85; Fig. 1F).

Stability of symptom ratings in healthy controls

Self-report PCSI total (Fig. 1A and Table 1) and subscale ratings (Supplementary Table S2) demonstrated a significant Time effect across retrospective SA and EC periods when limited to HC only. With the exception of the fatigue subscale, significantly higher retrospective ratings were observed relative to SA/EC (R>SA≈EC; all p's < 0.05) assessments, indicating a systematic reporting bias. Secondary measures (Supplementary Table S3) also indicated significant Time effects (all p's < 0.01). Anxiety and depression ratings indicated a similar pattern (R>SA≈EC; all p's < 0.05), whereas headache ratings were significantly elevated at retrospective and EC relative to SA assessments. Finally, sleep disturbances were elevated at EC relative to both retrospective and SA assessments.

For the parent-report PCSI, total score (Fig. 1B and Table 1) and all subscale ratings (Supplementary Table S2) with the exception of cognition ($p = 0.30$) exhibited main effects of Time (all p's < 0.05). Significantly elevated retrospective ratings were observed relative to SA/EC (R>SA≈EC; all p's < 0.05) for the PCSI total and physical scales. Emotional and fatigue subscales exhibited higher retrospective ratings relative to SA (all p's < 0.05) but not EC (p's > 0.05), with no differences between SA and EC ratings (p's > 0.10).

Inter-rater reliability

Within HC, inter-rater reliability (self vs. parent; Table 2) for PCSI total rating was poor for all (retrospective [ICC = 0.39]; SA [ICC = 0.24]; EC [ICC = 0.38]) assessments. Inter-rater reliability of subscales ranged from poor (majority of subscales) to good (physical subscale) across assessment periods. Within the pmTBI patient group, inter-rater reliability was poor for PCSI total retrospective ratings (ICC = 0.13), with correspondence increasing to good at SA (ICC = 0.70) and fair at EC (ICC = 0.55). A similar pattern of poor to fair reliability was observed for retrospective ratings across all PCSI subscales, with fair to good reliability at SA and EC periods for pmTBI patients.

Test-retest reliability of symptom measures

Within HC, test-retest reliability (Table 2) of self-report PCSI total ratings indicated excellent reliability between retrospective and SA (ICC = 0.85), fair reliability between retrospective and EC

TABLE 1. SYMPTOM REPORT RATINGS

Symptom measures	R HC	R pmTBI	SA HC (117)	SA pmTBI (162)	EC HC (94)	EC pmTBI (119)	HC stability
PCSI[a]	5(1–12)	8(1–27)	2(0–9)	20(7–50)	3(1–9)	9(1–26)	R>SA≈EC
PCSI (Parent)[a]	3(0–9)	3(1–12)	1(0–4)	13.5(6–36.5)	2(0–6.5)	6(1–13.5)	R>SA≈EC
PROMIS Anxiety[a]	2.5(1–6)	4(0–8)	1(0–3)	4(0–9)	1(0–4.5)	3(0–7)	R>SA≈EC
PROMIS Depression[a]	1(0–5)	3(0–9)	0(0–3)	2(0–11)	1(0–3)	2(0–7)	R>SA≈EC
PROMIS Sleep[a]	14(12–17.5)	18(14–23)	13(11–17)	20.5(15–25)	15(12–18.5)	20(14–25)	EC>R≈SA
HIT-6	44(38–51)	51(44–59)	40(36–47)	52.5(45–60)	42(38–47.5)	48(42–58)	R≈EC>SA

[a]Denotes significant group differences.

Data are formatted at median (interquartile range).

EC, early chronic; HC, healthy control; HIT-6, Headache Impact Test; PCSI, Post-Concussion Symptom Inventory; pmTBI, pediatric mild traumatic brain injury; PROMIS, Patient Reported Outcomes Measurement Information Systems; R, retrospective; SA, subacute.
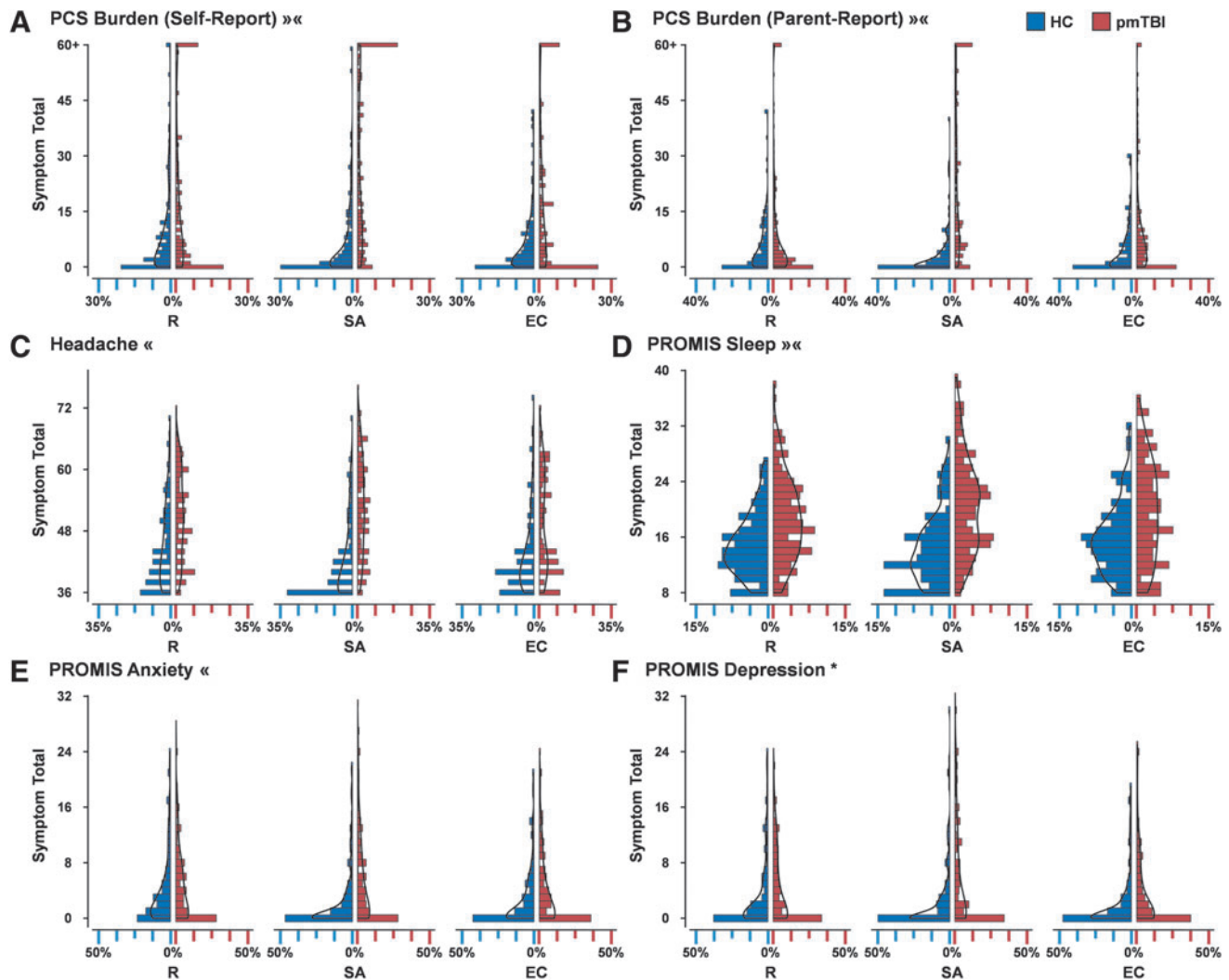
FIG. 1. Distributions of clinical measures. Histograms (bin = 1) represent the percentage (X axis) of healthy controls (HC; blue shading) and pediatric patients with mild traumatic brain injury (pmTBI; red shading) with various total symptom scores (Y axis) across retrospective (R), subacute (SA), and early chronic (EC) assessments. Density plots (solid black line) are overlaid on all histograms. The X axis is scaled to approximate the maximum percentage value among the six distributions for each symptom inventory. The Post-Concussion Symptom Inventory (PCSI) includes both self- **(A)** and parent-report **(B)** versions, whereas the Headache Impact Test (HIT-6) **(C)** and Patient Reported Outcomes Measurement Information System (PROMIS) measures for sleep **(D)**, anxiety **(E)**, and depression **(F)** are child self-report only. Results from ANCOVA models, using R ratings as a covariate, are denoted for significant Group effects (*) and Group×Time interactions («= group differences at SA only; »«= differences in magnitude across SA and EC periods).

(ICC = 0.50), and good reliability between SA and EC (ICC = 0.61) assessments. Subscales of the PCSI demonstrated similar or lower reliability. For the parent-report on PCSI, total ratings demonstrated good reliability (ICC = 0.67) between retrospective and SA and fair reliability between both retrospective and EC (ICC = 0.50) and SA and EC (ICC = 0.55) assessments. Reliability of PCSI subscales was similarly variable between assessment periods, but were generally lower relative to EC.

Test-retest reliability was fair between retrospective and both SA (ICC = 0.56) and EC (ICC = 0.58) assessments for self-report PCSI total ratings in pmTBI patients, but poor between SA and EC (ICC = 0.33) assessments, whereas subscale ratings ranged from poor to good (0.22–0.72). The decrement in reliability was notably greater for parent-reported PCSI relative to self-report, with poor reliability observed across all assessment periods (all ICC's ≤ 0.28), and subscale ratings ranging from poor to fair (0.09–0.50).

Test-retest reliability for self-report secondary measures ranged from fair to excellent in HC and fair to good in pmTBI patients (Supplementary Table S2).

False-positive rates and inter-method agreement of classification methods

Simple Change methods significantly misclassified HC as being symptomatic based on self-report at both SA (Fig. 2A; ICD-10: Z = 7.28; $p < 0.001$; Smyth: Z = 7.70; $p < 0.001$) and EC (Fig. 2B; ICD-10: Z = 13.40; $p < 0.001$; Smyth: Z = 12.45; $p < 0.001$) assessments, whereas the misclassification rate was significantly underestimated with RCI at the SA assessment (Z = 2.48; $p = 0.01$). For parent-report PCSI ratings, the Simple Change methods exceeded the false-positive misclassification rate only for EC assessment (Fig. 2B; ICD-10: Z = 9.10; $p < 0.001$; Smyth: Z = 10.57; $p < 0.001$).

TABLE 2. RELIABILITY COEFFICIENTS FOR POST-CONCUSSIVE SYMPTOMS

	Inter-Rater reliability						Test-Retest reliability: Child						Test-Retest reliability: Parent					
	R		SA		EC		R to SA		R to EC		SA to EC		R to SA		R to EC		SA to EC	
Measures	HC	pmTBI	HC	pmTBI	HC	pmTBI	HC	pmTBI	HC	pmTBI	HC	pmTBI	HC	pmTBI	HC	pmTBI	HC	pmTBI
PCSI																		
Total	0.39	0.13	0.24	0.70	0.38	0.55	0.85	0.56	0.50	0.58	0.61	0.33	0.67	0.28	0.50	0.17	0.55	0.24
Physical	0.64	0.13	0.36	0.75	0.50	0.46	0.75	0.54	0.28	0.49	0.52	0.23	0.78	0.21	0.62	0.23	0.64	0.23
Cognitive	0.30	0.04	0.32	0.52	0.16	0.37	0.83	0.51	0.38	0.47	0.43	0.27	0.78	0.29	0.44	0.21	0.43	0.24
Emotional	0.33	0.42	0.06	0.65	0.39	0.61	0.82	0.72	0.59	0.69	0.66	0.57	0.50	0.50	0.34	0.21	0.45	0.40
Fatigue	0.26	0.03	0.47	0.72	0.32	0.63	0.64	0.43	0.37	0.37	0.55	0.22	0.68	0.20	0.23	0.09	0.31	0.10

EC, early chronic; HC, healthy control; ICC, intraclass correlation coefficient; PCSI, Post-Concussion Symptom Inventory; pmTBI, pediatric mild traumatic brain injury; R, retrospective; SA, subacute.

In contrast, Standardized Change methods were within expected false-positive rates (all p's > 0.05) for self-report and parental ratings at both assessment periods.

Surprisingly, there was a 49% spread across methods for identifying pmTBI patients who would be diagnosed with PPCS at the EC visit (Fig. 3). Inter-method agreement for self-report was excellent for HC at the SA assessment (0.75–0.96), and excellent within Standardized Change methods at the EC assessment (0.95–0.99). Inter-method agreement across Simple Change and Standardized Change methods was notably lower at EC (0.53–0.65). Agreement across SA and EC assessments was excellent (0.89–0.98) within the Standardized Change methods, whereas poor to excellent agreement was observed for both Simple Change methods (0.38–0.75). A wider range of agreement was observed within self-report pmTBI ratings at SA (0.37–0.84) and EC (0.38–0.88) assessments, with higher agreement occurring within rather than between the two general method types. As expected due to recovery, agreement was poor to fair (0.12–0.59) for all methods across SA and EC assessments in pmTBI patients.

Inter-method agreement for parent-report ratings followed the same overall pattern as self-report ratings for both HC and pmTBI patients. However, inter-method agreement was generally higher for parent-report ratings with the exception of between SA and EC assessments for pmTBI patients (Fig. 3B, lower left quadrant).

Discussion

The primary aim of the current study was to investigate agreement between six different methods for defining significant symptom burden in both pmTBI patients and HC. Standardized PPCS methods greatly reduced misclassification rates within HC relative to Simple Change methods.[6,13,14] Importantly, misclassification rates among HC were relatively high (up to 35%) for both self and parental ratings when using the Simple Change methods that are employed in the majority of clinical and some research settings. These findings are not completely unexpected given the non-specific nature of PCS, and the number of typically developing adolescents who experience these symptoms.[9] The higher prevalence of misclassification rates at the EC visit may also be related to our use of retrospective ratings obtained at a single time-point closer to injury (i.e., approximately 1 week) to reduce recall error[10] rather than repeating retrospective symptom assessment at 4 months post-injury. Importantly, the current study defined misclassification of HC as above 5% and thus was more heavily weighted toward specificity rather than sensitivity of PPCS detection in pmTBI.

Inter-method agreement (e.g., PPCS vs. not PPCS) varied by method category (Standardized > Simple), assessment period (SA > EC), group (HC > pmTBI patients), and rater (parent > self). Excellent agreement was observed for both self-report and parent-report across methods at the SA assessment for HC, likely driven by our acquisition of retrospective and SA ratings on the same day. Similarly, excellent test-retest reliability was observed between retrospective and SA ratings for HC, suggesting the correspondence between various psychometric properties of an instrument. Conversely, test-retest reliability between retrospective/SA and EC ratings was only fair to good for HC. Inter-method agreement was also lower for HC at EC assessment, particularly for Simple Change methods. Previous studies report much higher test-retest reliability across 1–6 week reporting windows,[23,24] suggesting that duration between assessment periods is a critical factor for determining various aspects of reliability. Finally, the test-retest reliabilities for
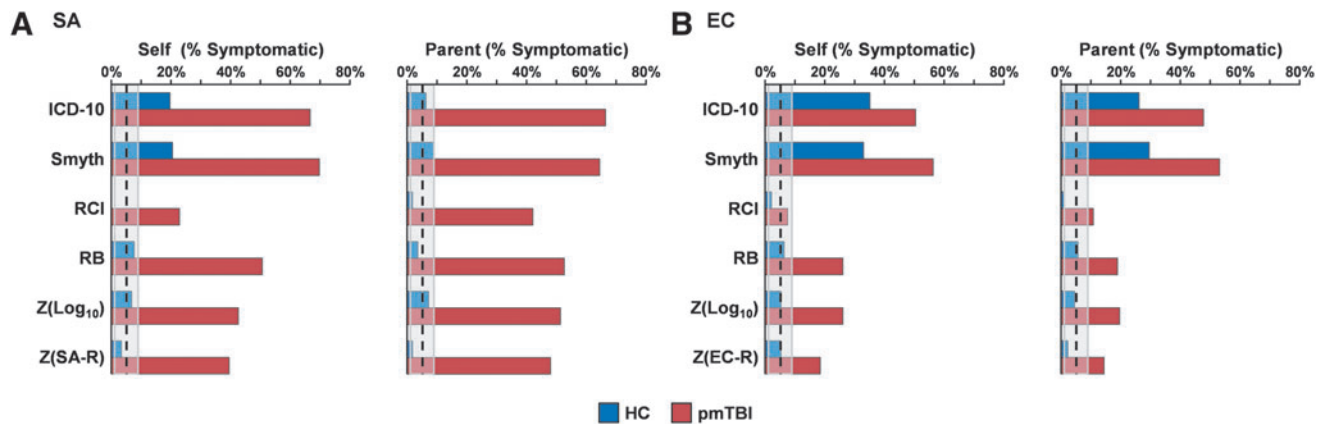
FIG. 2. Percent symptomatic rates. Percent symptomatic rates are displayed for healthy controls (HC; blue shading) and pediatric patients with mild traumatic brain injury (pmTBI; red shading) at the sub-acute (SA) **(A)** and early chronic (EC) **(B)** assessments. The X axis denotes the percentage of children classified as symptomatic based on either self or parental ratings from the Post-Concussion Symptom Inventory (PCSI) for six different methods (ICD-10: International Statistical Classification of Diseases and Related Health Problems, 10th revision; RCI: reliable change index; RB: regression-based). Retrospective ratings (R) were collected only at the SA period. The expected misclassification rate was operationally defined as 5%.

more specific symptom inventories (depression, anxiety, pain, etc.) and PCSI subscales were similar or lower among HC relative to total PCS burden, suggesting that the heterogeneous nature of PCSI inventories does not adversely affect reliability relative to more homogenous measures (i.e., all questions about depression).

Inter-method agreement for pmTBI patients across both self-report and parent-report classification was highly variable at both SA (full range across self-report and parent-report = 0.37–0.88) and EC (range = 0.38–0.95) assessments. In general, agreement for pmTBI patients was more variable between Simple versus Standardized Change methods rather than within each method. Importantly, a large spread was observed between methods in terms of the proportion of pmTBI patients "diagnosed" with PPCS at

4 months for both self (48.7% range; Fig. 2B) and parental (42.3% range) ratings. In contrast to current findings, a previous study comparing PPCS classification agreement among Simple Change methods relative to clinician judgement indicated superior performance for the ICD-10 method.[8] However, self-reported or parent-reported symptoms are typically the primary source for clinical determination of PPCS, rendering this validation approach somewhat tautological. The large variation in PPCS determination is similar to a recent ED study ($n = 11,907$) that indicated large variations (7.1%–98.7%) in the number of children who would be diagnosed with pmTBI across 17 different published definitions.[32] Although no true gold standard exists, current misclassification rates and classification spread among PPCS definitions suggests
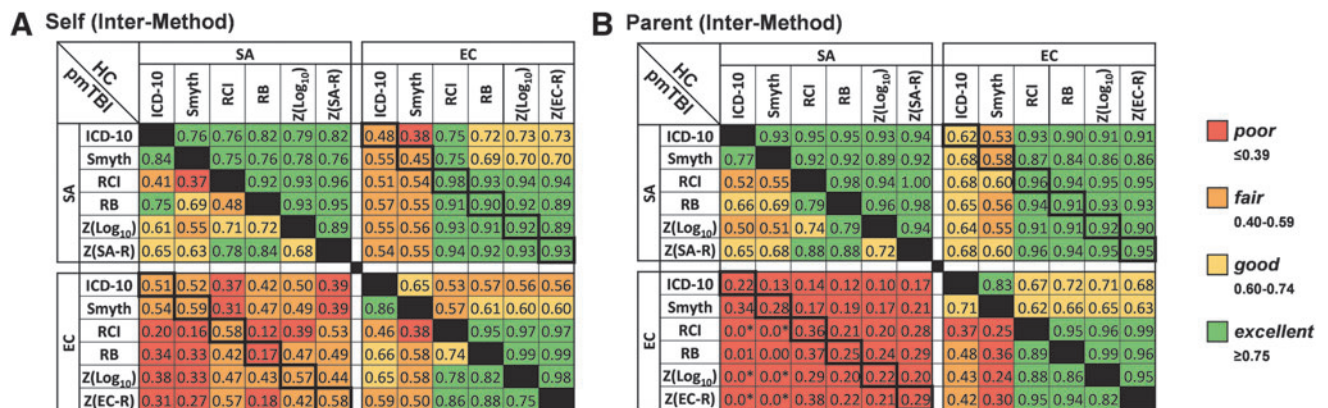

FIG. 3. Inter-method agreement. Inter-method agreement (poor: red; fair: orange; good: yellow; excellent: green) for binary classification (symptomatic vs. non-symptomatic/recovered) of each participant is presented in **(A)** (self rating) and **(B)** (parental rating). Data were derived from the Post-Concussion Symptom Inventory (PCSI) for six different methods (ICD-10: *International Statistical Classification of Diseases and Related Health Problems*, 10th revision; RCI: reliable change index; RB: regression-based). The upper left and lower right quadrants within each panel respectively denote agreement within the subacute (SA) or early chronic (EC) period, with healthy control (HC) data presented above the diagonal in these quadrants (pediatric mild traumatic brain injury [pmTBI] data below diagonal). In contrast, inter-method agreement between the SA and the EC assessment periods are presented in the upper right quadrant for HC or the lower left quadrant for pmTBI. Thick black rectangular outlines in the upper right or lower left panels denote self-correlation of methods across SA and EC periods. Off-diagonal elements are not equal in this matrix as a result of comparison from SA to EC versus EC to SA assessments. All negative agreement values were replaced with a lower limit of 0 and denoted with an asterisk (*).

that previous studies relying on Simple Change methods may overestimate the prevalence of PPCS among patient groups.

As would be expected due to changing symptoms following injury, both test-retest reliability and inter-method agreement (lower left quadrant, Fig. 3A,B) were generally lower for pmTBI patients across SA and EC assessments. Interestingly, although there was poor inter-rater reliability between self- and parental PCSI ratings across all assessment periods for HC, a noticeable improvement in parent/self-rating agreement occurred from the retrospective to the SA assessment in pmTBI patients, which then decreased again at the EC visit. Thus, increased day-to-day symptoms and/or heightened awareness (increased caregiver/child interactions) may improve inter-rater reliability to an acceptable level for proxy reporting in acutely injured patients. In contrast, parent/child agreement about retrospective, more chronic, or typical symptom (HC) burden is limited, potentially as a result of the statistical properties of limited symptom range. The poor reliability between child and parental estimates of symptom burden is particularly problematic given recent recommendations to use parental ratings in the calculation of risk scores for determining PPCS during more chronic injury stages.[6]

As alluded to in the previous section, stability of symptom burden ratings was assessed through a variety of different methods. Critically, although test-retest reliability was excellent for SA versus retrospective ratings in HC, retrospective ratings were significantly and consistently increased (i.e., more symptoms) relative to SA/EC ratings across most symptom inventories (PPCS, depression, anxiety, etc.), potentially reflecting a test-retest attenuation effect.[33] In contrast to current findings (pmTBI patients > HC), previous adult studies reported decreased retrospective ratings for patients (mTBI < HC), further suggesting this as a contributing factor for "good old days" biases.[15] Collectively, current and previous findings suggest multiple potential confounding effects involved in the use of retrospective ratings as a benchmark for estimating symptom burden.

In conclusion, many pre- and peri-injury factors have been associated with PPCS following mTBI, including acute symptom burden, adolescence at time of injury, female sex, history of psychiatric illness, history of prior concussion, and parental distress.[6,7,34] Both the ICD-10 and recent expert consensus panels[35] suggest a 30-day marker for defining prolonged symptoms in children. However, current and previous findings[9] question the specificity and utility of Simple Change scores. Given the poor correspondence between self-report and parent-report, as well as the possibility of elevated retrospective ratings in adolescents, current results suggest that the use of Standardized Change methods that either do not utilize retrospective ratings or statistically account for this bias (e.g., regression-based) represents the best available scientific and clinical option.

Previous[32] findings suggest large differences in who would be diagnosed with pmTBI, whereas current findings indicate considerable differences in who would be classified as "recovered." These findings collectively present considerable challenges for increasing the reliability of the pmTBI field as a whole. Although high-level evidence for treatments to shorten PPCS duration is limited,[35,36] PPCS can be managed effectively.[37] However, prior to the large-scale initiation of treatment trials, the field must first come to a consensus about what constitutes PPCS based on symptom duration and symptom burden. The use of different criteria across studies will continue to result in variable estimates of PPCS incidence, and will provide a large barrier for the care of children with concussion.[38,39]

Author Disclosure Statement

No competing financial interests exist.

Supplementary Material

Supplementary Appendix
Supplementary Table S1
Supplementary Table S2
Supplementary Table S3
Supplementary References

References

1. Mayer, A.R., Quinn, D.K., and Master, C.L. (2017). The spectrum of mild traumatic brain injury: a review. Neurology 89, 623–632.
2. Ontario Neurotrauma Foundation. (2013) *Guidelines for Concussion/Mild Traumatic Brain Injury and Persistent Symptoms*, 2nd ed. Ontario Neurotrauma Foundation: Toronto.
3. Management of Concussion-mild Traumatic Brain Injury Working Group. (2016). *VA/DoD Clinical Practice Guideline for Management of Concussion/Mild Traumatic Brain Injury*, 2nd ed. Department of Veterans Affairs and Department of Defense: Washington, DC.
4. Centers for Disease Control and Prevention. (2018). *Report to Congress: The Management of Traumatic Brain Injury in Children*. National Center for Injury Prevention and Control, Division of Unintentional Injury Prevention, Atlanta, GA.
5. Taylor, C.A., Bell, J.M., Breiding, M.J., and Xu, L. (2017). Traumatic brain injury-related emergency department visits, hospitalizations, and deaths - United States, 2007 and 2013. MMWR Surveill. Summ. 66, 1–16.
6. Zemek, R., Barrowman, N., Freedman, S.B., Gravel, J., Gagnon, I., McGahern, C., Aglipay, M., Sangha, G., Boutis, K., Beer, D., Craig, W., Burns, E., Farion, K.J., Mikrogianakis, A., Barlow, K., Dubrovsky, A.S., Meeuwisse, W., Gioia, G., Meehan, W.P., III, Beauchamp, M.H., Kamil, Y., Grool, A.M., Hoshizaki, B., Anderson, P., Brooks, B.L., Yeates, K.O., Vassilyadi, M., Klassen, T., Keightley, M., Richer, L., DeMatteo, C., and Osmond, M.H. (2016). Clinical risk score for persistent postconcussion symptoms among children with acute concussion in the ED. JAMA 315, 1014–1025.
7. Quinn, D.K., Mayer, A.R., Master, C.L., and Fann, J. (2018). Prolonged Postconcussive Symptoms. American Journal of Psychiatry 175, 103–111.
8. Hearps, S.J., Takagi, M., Babl, F.E., Bressan, S., Truss, K., Davis, G.A., Godfrey, C., Clarke, C., Doyle, M., Rausa, V., Dunne, K., and Anderson, V. (2017). Validation of a score to determine time to postconcussive recovery. Pediatrics 139, e20162003.
9. Iverson, G.L., Silverbergm N.D., Mannixm R., Maxwellm B.A., Atkinsm J.E., Zafonte, R., and Berkner, P/D. (2015). Factors associated with concussion-like symptom reporting in high school athletes. JAMA Pediatr. 169, 1132–1140.
10. Lange, R.T., Iverson, G.L., and Rose, A. (2010). Post-concussion symptom reporting and the "good-old-days" bias following mild traumatic brain injury. Arch. Clin. Neuropsychol. 25, 442–450.
11. Bressan, S., Takagi, M., Anderson, V., Davis, G.A., Oakley, E., Dunne, K., Clarke, C., Doyle, M., Hearps, S., Ignjatovic, V., Seal, M., and Babl, F.E. (2016). Protocol for a prospective, longitudinal, cohort study of postconcussive symptoms in children: the Take C.A.Re (Concussion Assessment and Recovery Research) study. BMJ Open 6, e009427.

12. Grubenhoff, J.A., Deakyne, S.J., Brou, L., Bajaj, L., Comstock, R.D., and Kirkwood, M.W. (2014). Acute concussion symptom severity and delayed symptom resolution. Pediatrics 134, 54–62.

13. Smyth, K., Sandhu, S.S., Crawford, S., Dewey, D., Parboosingh, J., and Barlow, K.M. (2014). The role of serotonin receptor alleles and environmental stressors in the development of post-concussive symptoms after pediatric mild traumatic brain injury. Dev. Med. Child Neurol. 56, 73–77.

14. Barlow, K.M., Crawford, S., Stevenson, A., Sandhu, S.S., Belanger, F., and Dewey, D. (2010). Epidemiology of postconcussion syndrome in pediatric mild traumatic brain injury. Pediatrics 126, e374–e381.

15. Gunstad, J., Suhr, J.A. (2001). "Expectation as etiology" versus "the good old days": postconcussion syndrome symptom reporting in athletes, headache sufferers, and depressed individuals. J. Int. Neuropsychol. Soc. 7, 323–333.

16. Yeates, K.O., Kaizar, E., Rusin, J., Bangert, B., Dietrich, A., Nuss, K., Wright, M., and Taylor, H.G. (2012). Reliable change in postconcussive symptoms and its functional consequences among children with mild traumatic brain injury. Arch. Pediatr. Adolesc. Med. 166, 615–622.

17. Jacobson, N.S., and Truax, P. (1991). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J. Consult. Clin. Psychol. 59, 12–19.

18. Iverson, G.L., Lovell, M.R., and Collins, M.W. (2003). Interpreting change on ImPACT following sport concussion. Clin. Neuropsychol. 17, 460–467.

19. Merritt, V.C., Bradson, M.L., Meyer, J.E., and Arnett, P.A. (2018). Evaluating the test-retest reliability of symptom indices associated with the ImPACT post-concussion symptom scale (PCSS). J. Clin. Exp. Neuropsychol. 40, 377–388.

20. Barker-Collo, S., Theadom, A., Jones, K., Ameratunga, S., Feigin, V., Starkey, N., Dudley, M., and Kahan, M. (2016). Reliable individual change in post concussive symptoms in the year following mild traumatic brain injury: data from the longitudinal, population-based Brain Injury Incidence and Outcomes New Zealand in the Community (Bionic) Study. JSM Burns Trauma 1, 1006.

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

22. Iverson, G.L., Gaetz, M., Lovell, M.R., and Collins, M.W. (2004). Cumulative effects of concussion in amateur athletes. Brain Inj. 18, 433–443.

23. Mailer, B.J., Valovich-McLeod, T.C., and Bay, R.C. (2008). Healthy youth are reliable in reporting symptoms on a graded symptom scale. J. Sport Rehabil. 17, 11–20.

24. Sady. M.D., Vaughan, C.G., and Gioia, G.A. (2014). Psychometric characteristics of the postconcussion symptom inventory in children and adolescents. Arch. Clin. Neuropsychol. 29, 348–363.

25. McCrory, P., Meeuwisse, W.H., Aubry, M., Cantu, B., Dvorak, J., Echemendia, R.J., Engebretsen, L., Johnston, K., Kutcher, J.S., Raftery, M., Sills, A., Benson, B.W., Davis, G.A., Ellenbogen, R.G., Guskiewicz, K., Herring, S.A., Iverson, G.L., Jordan, B.D., Kissick, J., McCrea, M., McIntosh, A.S., Maddocks, D., Makdissi, M., Purcell, L., Putukian, M., Schneider, K., Tator, C.H., and Turner, M. (2013). Consensus statement on concussion in sport the 4th International Conference on Concussion in Sport held in Zurich, November 2012. Br. J. Sports Med. 47, 250–258.

26. Chelune, G.J., Naugle, R.I., Lüders, H., Sedlak, J., and Awad, I.A. (1993). Individual change after epilepsy surgery: practice effects and base-rate information. Neuropsychol. 7, 41.

27. Mayer, A.R., Bedrick, E.J., Ling, J.M., Toulouse, T., and Dodd, A. (2014). Methods for identifying subject-specific abnormalities in neuroimaging data. Hum. Brain Mapp. 35, 5457–5470.

28. Shrout, P.E., and Fleiss, J.L. (1979). Intraclass correlations: uses in assessing rater reliability. Psychol. Bull. 86, 420–428.

29. Gwet, K.L. (2008). Computing inter-rater reliability and its variance in the presence of high agreement. Br. J. Math Stat. Psychol. 61, 29–48.

30. Cicchetti, DV. (2001). The precision of reliability and validity estimates re-visited: distinguishing between clinical and statistical significance of sample size requirements. J. Clin. Ex.p Neuropsychol. 23, 695–700.

31. Clopper, C.J., and Pearson, E.S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika 26, 404–413.

32. Crowe, L.M., Hearps, S., Anderson, V., Borland, M.L., Phillips, N., Kochar, A., Dalton, S., Cheek, J.A., Gilhotra, Y., Furyk, J., Neutze, J., Lyttle, M.D., Bressan, S., Donath, S., Molesworth, C., Oakley, E., Dalziel, S.R., and Babl, F.E. (2018). Investigating the variability in mild traumatic brain injury definitions: a prospective cohort study. Arch. Phys. Med. Rehabil. 99, 1360–1369.

33. Achenbach, T.M., and Rescorla, L.A. Manual for the ASEBA Adult Forms and Profiles. Burlington, VT: University of Vermont, Research Center for Children, Youth and Families.

34. Rausa, V.C., Anderson, V., Babl, F.E., and Takagi, M. (2018). Predicting concussion recovery in children and adolescents in the emergency department. Curr. Neurol. Neurosci. Rep. 18, 78.

35. McCrory, P., Meeuwisse, W., Dvorak, J., Aubry, M., Bailes, J., Broglio, S., Cantu, R.C., Cassidy, D., Echemendia, R.J., Castellani, R.J., Davis, G.A., Ellenbogen, R., Emery, C., Engebretsen, L., Feddermann-Demont, N., Giza, C.C., Guskiewicz, K.M., Herring, S., Iverson, G.L., Johnston, K.M., Kissick, J., Kutcher, J., Leddy, J.J., Maddocks, D., Makdissi, M., Manley, G.T., McCrea, M., Meehan, W.P., Nagahiro, S., Patricios, J., Putukian, M., Schneider, K.J., Sills, A., Tator, C.H., Turner, M., and Vos, PE. (2017). Consensus statement on concussion in sport - the 5(th) International Conference on Concussion in Sport held in Berlin, October 2016. Br. J. Sports Med. 51, 838–847.

36. Lumba-Brown, A., Yeates, K.O., Sarmiento, K., Breiding, M.J., Haegerich, T.M., Gioia, G.A., Turner, M., Benzel, E.C., Suskauer, S.J., Giza, C.C., Joseph, M., Broomand, C, Weissman, B., Gordon, W., Wright, D.W., Moser, R.S., McAvoy, K., Ewing-Cobbs, L., Duhaime, A.C., Putukian, M., Holshouser, B., Paulk, D., Wade, S.L., Herring, S.A., Halstead, M., Keenan, H.T., Choe, M., Christian, C.W., Guskiewicz, K., Raksin, P.B., Gregory, A., Mucha, A., Taylor, H.G., Callahan, J.M., DeWitt, J., Collins, M.W., Kirkwood, M.W., Ragheb, J., Ellenbogen, R.G., Spinks, T.J., Ganiats, T.G., Sabelhaus, L.J., Altenhofen, K., Hoffman, R., Getchius, T., Gronseth, G., Donnell. Z,. O'Connor, R.E., and Timmons, S.D. (2018). Centers for Disease Control and Prevention Guideline on the Diagnosis and Management of Mild Traumatic Brain Injury among Children. JAMA Pediatr. 172, e182853.

37. Schneider, K.J., Leddy, J.J., Guskiewicz, K.M., Seifert, T., McCrea, M., Silverberg, N.D., Feddermann-Demont, N., Iverson, G.L., Hayden, A., and Makdissi, M. (2017). Rest and treatment/rehabilitation following sport-related concussion: a systematic review. Br. J. Sports Med. 51, 930–934.

38. McCauley, S.R., Boake, C., Pedroza, C., Brown, S.A., Levin, H.S., Goodman, H.S., and Merritt. S.G. (2005). Postconcussional disorder: are the DSM-IV criteria an improvement over the ICD-10? J. Nerv. Ment. Dis. 193, 540–550.

39. Rose, S.C., Fischer, A.N., and Heyer, G.L. (2015). How long is too long? The lack of consensus regarding the post-concussion syndrome diagnosis. Brain Inj. 29, 798–803.

Address correspondence to:
Andrew R. Mayer, PhD
The Mind Research Network
Pete and Nancy Domenici Hall
1101 Yale Boulevard NE
Albuquerque, NM 87106
USA

E-mail: amayer@mrn.org