A Data-driven Approach to Unlikely, Possible, Probable, and Definite Acute Concussion Assessment

Gian-Gabriel P. Garcia, MS[1]; Mariel S. Lavieri, PhD[1]; Ruiwei Jiang, PhD[1]; Thomas W. McAllister, MD[2]; Michael A. McCrea, PhD[3]; Steven P. Broglio, PhD[4]; CARE Consortium Investigators

[1]Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, Michigan, USA

[2]Department of Psychiatry, Indiana University School of Medicine, Indianapolis, IN, USA

[3]Departments of Neurosurgery and Neurology, Medical College of Wisconsin, Milwaukee, WI, USA

[4]School of Kinesiology, University of Michigan, Ann Arbor, Michigan, USA

The Version of Record of this manuscript has been published and is available in *Journal of Neurotrauma*, 6 May 2019. https://www.liebertpub.com/doi/10.1089/neu.2018.6098

Abstract (250 words): Kutcher and Giza suggested incorporating levels of certainty in concussion diagnosis decisions. These guidelines were based on clinical experience rather than objective data. Therefore, we combined data-driven optimization with predictive modeling to identify which athletes are Unlikely to have concussion and classify remaining athletes as Possible, Probable, or Definite concussion by diagnostic certainty. We developed and validated our framework using data from the Concussion Assessment, Research, and Education (CARE) Consortium. Acute concussions consisted of assessments at <6 hours (n=1085) and 24-48 hours post-injury (n=1413). Normal performances consisted of assessments at baseline (n=1635) and the time of unrestricted return-to-play (n=1345). We evaluated the distribution of acute concussions and normal performances across risk categories and identified interclass and intraclass differences in demographics, time-of-injury characteristics, the Standard Assessment of Concussion (SAC), Sport Concussion Assessment Tool (SCAT) symptom assessments, and Balance Error Scoring System (BESS). Our algorithm accurately classified concussions as Probable or Definite (sensitivity=91.07-97.40%). Definite and Probable concussions had higher SCAT symptom scores compared to Unlikely and Possible concussions ($p<0.05$). Definite concussions had lower SAC and higher BESS scores ($p<0.05$). Baseline to post-injury change scores for the SAC, SCAT symptoms, and BESS were significantly different between acute Possible or Probable concussions and normal performances ($p<0.05$). There were no consistent patterns in demographics across risk categories, although a greater proportion of concussions classified as Unlikely were reported immediately compared to Definite concussions ($p<0.05$). While clinical interpretation is still needed, our data-driven approach to concussion risk stratification provides a promising step towards evidence-based concussion assessment.

Key words: Possible, Probable, and Definite concussion; acute concussion assessment; risk stratification

INTRODUCTION

Concussion, a type of traumatic brain injury, is an important public health issue that has been associated with potential long-term health consequences.[1] Accurate diagnosis and proper post-injury concussion management of concussion are pragmatic steps for mitigating possible consequences.[2] However, the clinical diagnosis of concussion is challenging for many reasons,[3,4] especially in the sporting environment which requires rapid injury assessment and injury management decisions.

Currently, no diagnostic marker or clinical assessment has been designed for perfectly identifying concussion. To this end, existing guidelines call for the evaluation of concussion along multiple domains, including symptom presentation, neurocognitive status, and a physical examination.[5–7] Previous studies analyzing such multidimensional testing batteries for acute concussion assessment have found that symptom evaluations were the most sensitive component of these batteries.[8–12] However, an over-reliance on symptom presentation for concussion diagnosis is troublesome for several reasons. First, symptoms may be under-reported or go completely unrecognized.[13–15] Furthermore, common symptoms (e.g., headache, dizziness, fatigue) which are indicative of concussion are not necessarily specific to concussion. Finally, rapid changes in symptom and neurocognitive presentation within the acute stages of concussion results in highly variable clinical presentation of concussion across multiple patients.[16,17]

To address these challenges, Kutcher and Giza recommended incorporating diagnostic certainty to the assessment of concussion.[4] That is, rather than a binary diagnosis paradigm (i.e., concussion or no concussion), Kutcher and Giza suggested that concussion diagnosis should be relayed across a spectrum of risk categories (e.g., Possible, Probable, and Definite concussion), with each category reflecting the degree to which a concussion diagnosis is certain. Similar risk-based categories have been used for classifying diagnosis decisions for other diseases, including multiple sclerosis,[18] Alzheimer's disease,[19] and diabetes.[20] Compared to traditional binary diagnosis, risk-based diagnosis frameworks account for the evolution of the injury over time and allows for more flexibility in the post-injury management of concussion. Specifically, incorporating certainty in the assessment of concussion can help to determine whether an athlete should be managed *as if* he or she has a concussion, ultimately improving the quality of patient care. However, the guidelines developed by Kutcher and Giza were based on clinical experience rather than objective data.

Therefore, the goal of our study is to create a data-driven modeling framework to identify concussed and non-concussed athletes who are *Unlikely* to have concussion and classify the remaining athletes as having *Possible, Probable,* or *Definite* concussion, with each category reflecting increasing diagnostic certainty. While experienced clinicians may be able to quickly synthesize the likelihood of concussion and ultimately identify a post-injury management plan for athletes, our data-driven framework provides a more objective approach which can ultimately benefit those clinicians who may be inexperienced in managing concussion. We then aim to validate our framework to identify how athletes are distributed across each risk classification and identify differences in demographics, time-of-injury characteristics, and standard assessment scores between athletes under each risk classification.

MATERIALS AND METHODS

STUDY POPULATION AND DESIGN

To develop our methodology for classifying athletes as Unlikely, Possible, Probable, or Definite concussion, we used data from the Concussion Assessment, Research, and Education (CARE) Consortium.[21] The CARE Consortium defines concussion as "a change in brain function following a force to the head, which may be accompanied by temporary loss of consciousness, but is identified in awake individuals with measures of neurologic and cognitive dysfunction."[22] These acquired data contain 33,271 player-seasons collected during the 2014-2018 academic years from 29 National College Athletic

Association (NCAA) universities and military service academies. Player-season data were collected across male (57.58%) and female (42.42%) participants from 27 sports, including 19.8% from football, 12.1% from cross country/track, and 9.6% from soccer. The data include 24,561 athletes with pre-season baseline evaluations and 1,950 concussions across 1,755 athletes. For student-athletes (hereafter, referred to as just athletes) who were diagnosed with concussion by the local institution's medical staff (e.g., team physicians and athletic trainers), additional post-injury data were collected within 6 hours of the injury (<6 hours), 24-48 hours post-injury (24-48 hours), when s/he was identified as asymptomatic, when cleared for unrestricted return-to-play (RTP), and 6 months post-RTP. We note that some athletes may not have completed a post-injury assessment at every timepoint. Therefore, there is some missingness in the data leading to unequal sample sizes across study cohorts. However, these imbalances are not to a degree which has significant effects on the methodologies employed in this research. All participants provided written consent which was approved by their local institutional review board and the US Army Human Research Protection Office.

SAMPLE SELECTION

In our analysis, we focused on the timepoints at baseline, <6 hours, 24-48 hours, and unrestricted RTP. We only included baseline data which could be matched with post-injury data. The assessments at <6 hours and 24-48 hours were denoted "acute concussion" and those from baseline and unrestricted RTP were denoted "normal performance". We consider those from the unrestricted RTP timepoint to demonstrate a normal performance because they have been cleared for RTP by each institution's local medical staff. We analyzed <6 hours and 24-48 hours separately.

STUDY VARIABLES

For each participant in the study data, we obtained demographic information along with raw scores on the Standard Assessment of Concussion (SAC), Standard Concussion Assessment Tool (SCAT) symptom survey, and the Balance Error Scoring System (BESS) at baseline. For those diagnosed with concussion, we obtained time-of-injury characteristics along with raw scores for SAC, SCAT symptoms, and BESS scores at each post-injury evaluation timepoint. We computed the change score for these athletes by subtracting the raw score at baseline from the raw score at each post-injury timepoint. A positive change score indicates an increase in the measure compared to baseline, whereas a negative change score indicates a decrease compared to baseline. We also filled missing data elements using multiple imputation by chained equations.[23] We describe our study variables in more detail below.

DEMOGRAPHIC INFORMATION

We aimed to identify differences in each risk classification by age, sex, and the number of previous concussions. Previous studies have suggested that younger athletes, females, and those with greater concussion history are at increased risk for concussion. [24–29]

TIME-OF-INJURY CHARACTERISTICS

In our analysis, we included whether the athlete experienced loss of consciousness (LOC), post-traumatic amnesia (PTA), retrograde amnesia (RGA), whether the athlete was removed from play immediately, and whether the injury was reported immediately, as these variables have been suggested to impact concussion risk.[30–33]

STANDARD ASSESSMENT OF CONCUSSION

The SAC is a neurocognitive assessment which measures orientation, immediate memory, concentration, and delayed recall.[34] In our analysis, we focused on the SAC total score and change score, both of which summarize the SAC assessment.

SPORT CONCUSSION ASSESSMENT TOOL SYMPTOM SURVEY

Symptom presentation has been shown, in numerous studies, to be highly associated with acute concussion.[8–12] The SCAT symptom evaluation includes 22 symptoms, each of which is rated on a scale of

0-6 based on severity.[35] In our analysis, we included the total symptom severity and the total number of symptoms in addition to their respective change scores.

BALANCE ERROR SCORING SYSTEM

The BESS is a physical examination which measures postural stability by assessing the number of "movement errors" committed by an athlete while attempting to hold different stances.[36] Balance has been noted to be affected by concussion and we included the BESS total score (across all 6 stances) and change score in our analysis.

DATA ANALYSIS

Our overall framework for classifying *Unlikely*, *Possible*, *Probable*, or *Definite* concussion is summarized in **Figure 1**. To create and evaluate our models, we divided our post-injury data into a training set and a validation set. The training set consisted of all data collected between January 23, 2014 and November 29, 2016 while the validation data consisted of all data collected after that date (i.e., November 30, 2016 to October 2, 2017). The CARE Consortium protocol for concussion diagnosis along with assessments performed at baseline and post-injury remained unchanged during this period and thus, rater drift was minimal, if any. We used our training set to develop the models to determine which athletes should be classified under each risk category. Then, we applied our models to the validation data to evaluate and analyze our framework. We describe each of the steps in our methodology in more detail below.

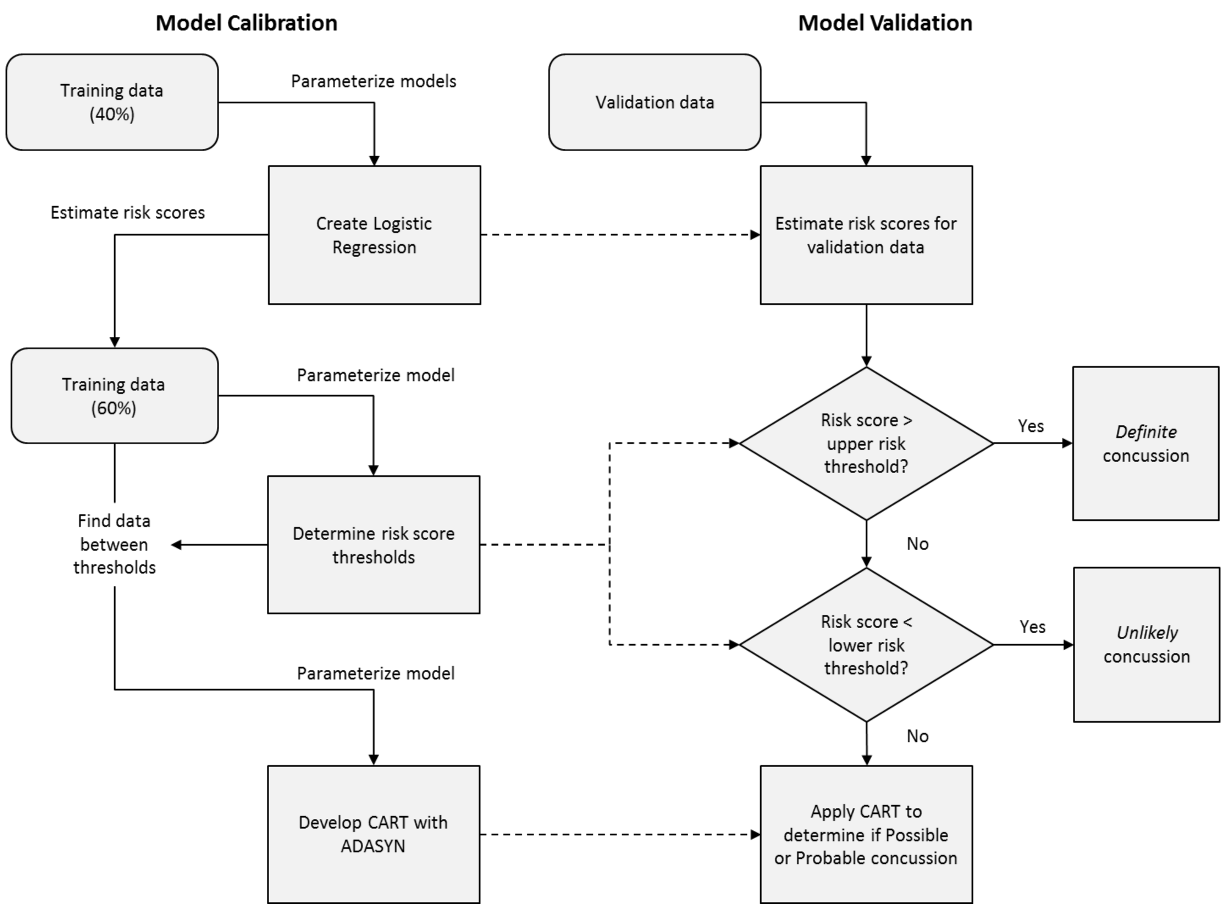

FIGURE 1. ILLUSTRATION OF METHODOLOGICAL FRAMEWORK FOR DEVELOPING DATA-DRIVEN MODELS WHICH WERE

MODEL CALIBRATION

Using a randomly chosen 40% of the training data, we created a logistic regression model for estimating risk scores. For any athlete, risk scores are a scalar between 0 and 1, where greater risk scores indicate higher likelihood of acute concussion. We used a previously published and validated multivariate logistic regression model (i.e., the raw score model) to estimate risk scores associated with athletes at <6 hours and 24-48 hours.[12] We used the raw score models since change scores could not be computed for baseline data and may not always be available for acute concussion assessments in clinical settings. Since time-of-injury characteristics were not available for baselines but were part of these previous models, we assumed in this logistic regression analysis only that for baseline data, injuries were reported immediately and that participants were removed from play immediately.

With the remaining 60% of the training data, we determined risk score thresholds to identify Unlikely and Definite concussions. We first applied our logistic regression models to this subset of training data to obtain risk scores for each athlete. Then, we used these risk scores as the input for a previously developed data-driven optimization algorithm to determine risk score thresholds.[*] This algorithm identifies an upper and lower risk threshold by maximizing sensitivity and specificity while limiting false-positive and false-negative rates. Athletes with risk scores below the lower threshold represent those whose concussion probability was low and thus would be identified as Unlikely concussion. Similarly, athletes with risk scores above the upper threshold are most likely to have concussion and would be classified as Definite concussion. In designing these thresholds, we favored higher sensitivity over lower false-positive rates and lower false-negative rates over higher specificity.

After determining the upper and lower risk score thresholds, we identified athletes in the training set with risk scores between Unlikely and Definite thresholds and used them to determine how athletes should be classified as Possible or Probable, as these cases could not be easily distinguished by our logistic regression model. Categorization of these cases was approached using a classification and regression tree (CART)[37] analysis. CART is a non-parametric statistical modeling technique which produces a decision tree for prediction and is capable of handling categorical variables and continuous variables. Compared to other predictive modeling methods (e.g., generalized linear models), CART is advantageous in its interpretability and ability to model highly non-linear relationships between variables. Due to the higher proportion of normal performances to acute concussions in our data, we applied Adaptive Synthetic Sampling (ADASYN) to mitigate data imbalance issues before creating a CART.[38] Additionally, for this CART, we restricted the resulting decision tree to include only variables which were available for all timepoints. That is, the resulting decision tree did not include time-of-injury

[*] Garcia, G.-G.P., Lavieri, M.S., Jiang, R., McCrea, M., McAllister, T.W., Broglio, S.P., and CARE Consortium Investigators. Data-driven stochastic optimization approaches to determine decision boundaries for medical diagnosis. Submitted for publication.

characteristics and change scores for the SAC, SCAT symptom assessments, and the BESS since they were not available for baseline data. Athletes who were predicted to be acute concussions by this CART were classified as Probable concussions while those who were predicted to be normal performances were classified as Possible concussions.

MODEL VALIDATION

To implement our models, we applied our logistic regression models to the validation data to obtain risk scores for each athlete. Then, we compared these risk scores to the upper and lower thresholds we generated using our optimization algorithm in the model calibration phase. Athletes with risk scores below the lower threshold were classified as Unlikely concussions, while athletes with risk scores above the upper threshold were classified as Definite concussions. We then applied our CART to any athlete with a risk score between these thresholds to classify them as Possible or Probable concussion.

MODEL EVALUATION

After implementing our models on the validation data, we performed additional analysis to evaluate the performance of our classification framework. The goals of this analysis were to (1) analyze how our models classified acute concussions and normal performance throughout each risk category and (2) identify interclass differences (i.e., across different risk classifications) and intraclass differences (i.e., within the same risk classification) in demographics, time-of-injury characteristics, and standard assessment scores for acute concussions and normal performances among the risk classifications.

To achieve the first goal, we determined the percentage of acute concussions and normal performances within each risk classification at both <6 hours and 24-48 hours. Ideally, data captured in the acute post-injury state should place the athlete in greater risk classifications (i.e., Definite or Probable), while data captured at baseline should place the athlete into lower risk classifications (i.e., Unlikely or Possible). We compared the distribution of acute concussions and normal performances using the Kolmogorov-

Smirnov test. A significant p-value for this test implies that the distribution of acute concussions and normal performances among the risk classifications is dissimilar.

Since our diagnosis scheme consisted of 4 risk categories instead of 2, we also computed a modified sensitivity and specificity. Our modified computation was founded in recommendations by Kutcher and Giza who indicated that Probable and Definite concussions should be managed as concussions, while Possible concussions should be managed based on clinical judgment. Furthermore, we assume that Unlikely concussions are managed as non-concussions. Therefore, we provide a sensitivity range where the lower bound reflects the proportion of acute concussions that are correctly classified as Probable and Definite and the upper bound reflects the proportion of acute concussions correctly classified as Possible, Probable and Definite. We also provide a range for specificity, where the lower bound reflects the case where no Possible concussions are treated as non-concussed and an upper bound which reflects the case where all Possible concussions are treated as non-concussed. In practice, the true sensitivity and specificity should fall between these bounds depending on how Possible concussions are managed.

To achieve the second goal, we first identified interclass differences in the study variables across each risk classification for acute concussions and normal performances using analysis of variance (ANOVA) tests with Tukey's post-hoc comparisons. For example, we determined if athletes with acute concussion who were classified as having a Probable concussion had any differences in SAC, SCAT symptoms, or BESS compared to acute concussions who were classified as Definite concussions. Next, using Student's t-test, we identified intraclass differences in the study variables between acute concussions and normal performances within each risk classification. All models were created and analyzed using Python 3.5.2 (Python Software Foundation, Beaverton, Oregon, USA).

RESULTS

In **Table 1**, we summarize the study data at each timepoint with respect to the study variables. Across all timepoints, there were significant differences between training and validation data in height (p=0.0082-0.047), weight (p=0.012-0.047), and number of previous concussions (p<0.001 for all). There were also significant differences in age at baseline (p=0.0012) and 24-48 hours (p=0.021) and the proportion of males at unrestricted RTP (p=0.013). Among post-injury assessments, there were significant differences between SAC raw scores at baseline (p=0.00085), <6 hours (p=0.038), and at unrestricted RTP (p=0.0038), SCAT total symptoms raw score at unrestricted RTP (p=0.027), and BESS raw score at <6 hours (p=0.048) and 24-48 hours (p=0.00098).

MULTIVARIATE LOGISTIC REGRESSION

The model variables and corresponding coefficient values for the multivariate logistic regression models at <6 hours and 24-48 hours are shown in **Table 2**. At <6 hours, all variables were significant except for whether the injury was reported immediately (p=0.16), SAC raw score (p=0.13), and BESS raw score (p=0.080). At 24-48 hours, all variables were significant except for SAC raw score (p=0.23) and BESS raw score (p=0.94).

CLASSIFYING UNLIKELY, POSSIBLE, PROBABLE, AND DEFINITE CONCUSSION

We obtained risk score thresholds after applying the training data to our optimization algorithm. At <6 hours, the lower threshold was 0.047 and the upper threshold was 0.33. At 24-48 hours, the lower threshold was 0.07 and the upper threshold was 0.46. The CART we developed for <6 hours and 24-48 hours are shown in **Figure 2**.

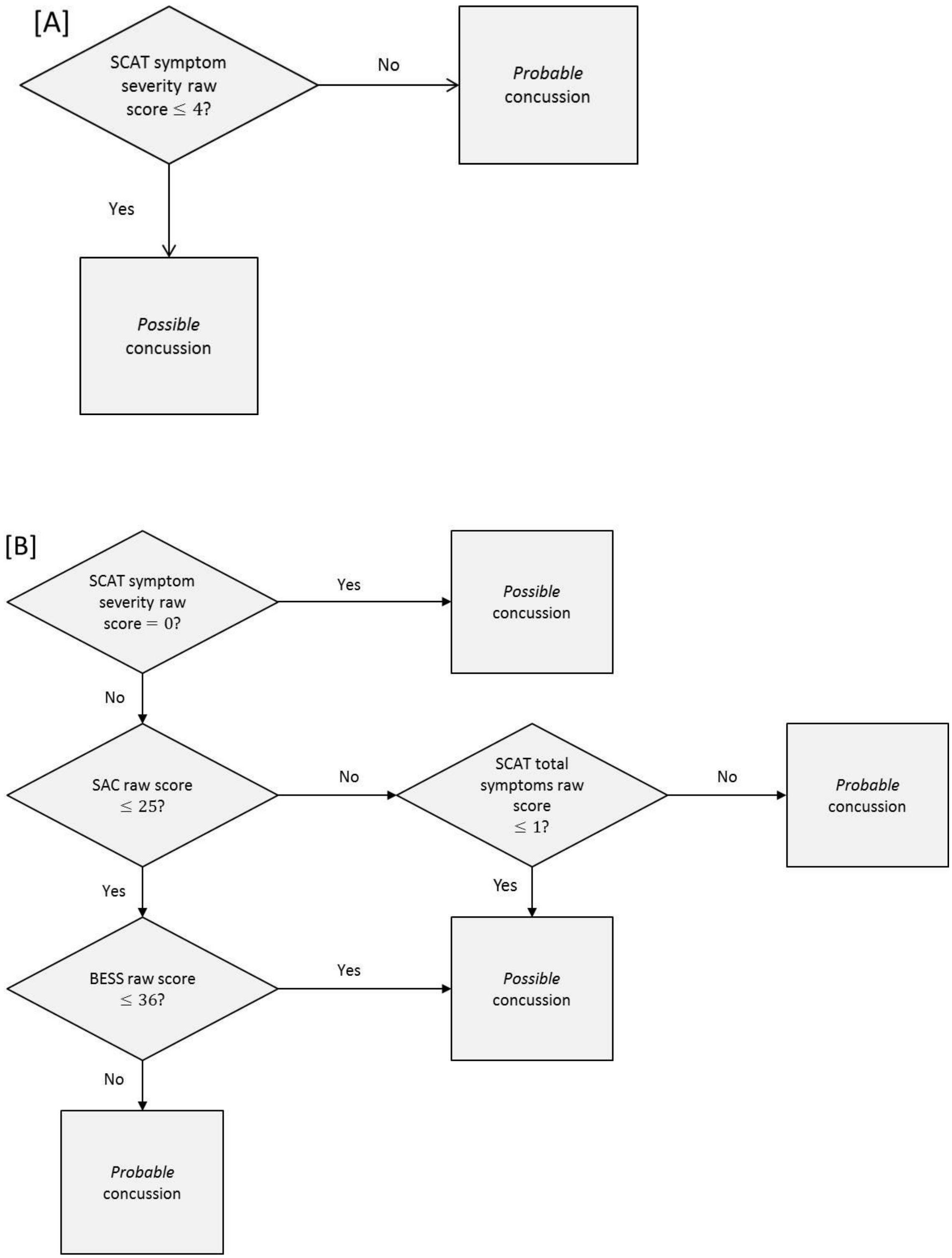

FIGURE 2. CLASSIFICATION TREE FOR DETERMINING POSSIBLE AND PROBABLE CONCUSSIONS AT [A] <6 HOURS AND [B] 24-48 HOURS POST-INJURY

We now provide an example to illustrate how these risk thresholds and CART can be used to determine whether an athlete should be classified as an Unlikely, Possible, Probable, or Definite concussion.

Consider a 19 year-old female athlete who is being assessed for acute concussion 24-48 hours after injury. She did not report the injury immediately and in her post-injury assessments, obtained total scores of 30 and 12 on the SAC and BESS, respectively. On the SCAT symptom assessment, she reported 4 total symptoms with a total severity of 6. Using the logistic regression model for 24-48 hours, her risk score is equal to 0.36. Since her risk score is less than the upper threshold of 0.46 and greater than the lower threshold of 0.07 at 24-48 hours, she is not classified as a Definite or Unlikely concussion. To determine if she is a Possible or Probable concussion, one would refer to the CART for 24-48 hours. Since her SCAT symptom severity raw score is not 0, her SAC raw score is greater than 25, and her SCAT total symptoms raw score is greater than 1, she would be classified as a Probable concussion.

To provide an additional example, consider a 21 year-old male athlete who was assessed for concussion within 6 hours of a suspected injury. His injury was not reported immediately and he was not removed from play immediately. His SAC raw score and BESS raw score were 24 and 12, respectively. He also reported 1 symptom with a severity of 1. Based on these values, his risk score is equal to 0.22. Since his risk estimate is between the lower and upper thresholds of 0.047 and 0.33 at <6 hours, respectively, then he must either be a Possible or Probable concussion. Since his SCAT symptom severity raw score is ≤4, the CART analysis at <6 hours would classify him as a Possible concussion.

DISTRIBUTION OF ACUTE CONCUSSIONS AND NORMAL PERFORMANCES

The distribution of acute concussions and normal performances within each risk classification is shown in **Table 3**. At <6 hours, 434 (80.52%) of acute concussions were classified as Definite concussion while only 14 (2.60%) were classified as Unlikely concussion. Among the remaining acute concussions, 31 (5.75%) were classified as Possible concussion and 60 (11.13%) were classified as Probable concussion.

When the <6 hours algorithm was applied to normal performance data (i.e., baseline and unrestricted RTPs), 696 (46.00%), 526 (34.77%), 189 (12.49%), and 102 (6.74%) were classified as Unlikely, Possible, Probable, and Definite concussion respectively. With the 24-48 hours algorithm, 522 (75.22%) acute concussions were classified as Definite concussion while 21 (3.03%) were classified as Unlikely concussion. There were 41 (5.91%) and 110 (15.85%) acute concussions classified as Possible and Probable concussion, respectively. Among the normal performances, 714 (47.19%), 397 (26.24%), 309 (20.42%), and 95 (6.15%) were classified as Unlikely, Possible, Probable, and Definite concussion, respectively. With both <6 hours and 24-48 hours algorithms, the distributions among risk classifications were different between acute concussions and normal performances based on the Kolmogorov-Smirnov test ($p<0.001$). Additionally, the distribution of baselines and unrestricted RTPs across the risk classifications were also significantly different at both <6 hours and 24-48 hours ($p<0.001$).

Using our modified calculation for sensitivity and specificity, we obtained a sensitivity range of 91.65-97.40% with the <6 hours algorithm and 91.06-97.00% with 24-48 hours algorithm. We also obtained a specificity range of 46.00-80.77% with the <6 hours algorithm and 47.19-73.43% with the 24-48 hours algorithm, respectively.

As an ancillary analysis, we performed our analysis without the unrestricted RTP data (data not shown). The resulting logistic regression model, risk score thresholds, and CART models led to a distribution with a sensitivity of 89.42%-98.52% and a specificity of 23.21%-71.60% at <6 hours. At 24-48 hours, the sensitivity and specificity ranged from 85.73%-95.67% and 41.52%-71.60%, respectively.

INTERCLASS DIFFERENCES

The interclass differences for acute concussions and normal performances are shown in **Table 4** and **Table 5**, respectively. Among acute concussions, all mean raw and change scores for SCAT symptom assessments at Unlikely and Possible concussions are significantly different from Definite concussion at both <6 hours and 24-48 hours (p<0.001 for all). Among the SAC and BESS at <6 hours, only the SAC change score is not significantly different between Definite and Unlikely concussions. In contrast, at 24-48 hours, only BESS raw score is significantly different between Definite and Unlikely concussions (p=0.021). Possible and Probable concussions are significantly different in SCAT total symptoms raw score at <6 hours and 24-48 hours (p=0.0027-0.0082). They are also significantly different in SAC change score (p=0.013), SAC raw score (p<0.001), and SCAT total symptoms change score (p=0.012) at 24-48 hours.

For normal performances, the mean raw scores for the SAC, SCAT symptom severity, SCAT total symptoms, and the BESS among Definite and Probable concussions are significantly different from Unlikely concussion (p<0.001 for all), except for SAC raw score at 24-48 hours. At <6 hours and 24-48 hours, Possible and Probable concussions were significantly different in SCAT symptom severity raw score and SCAT total symptoms raw score (p<0.001 for all). At 24-48 hours, Possible and Probable concussions are also significantly different in SAC raw score and BESS raw score (p<0.001 for all).

INTRACLASS DIFFERENCES

The intraclass differences are highlighted in **Table 3** and **Table 4**. Among those classified as Possible concussions, acute concussions and normal performances are significantly different in SCAT symptom severity (p<0.001 at <6 hours, p=0.016 at 24-48 hours) and SCAT total symptoms raw score (p<0.001 at <6 hours, p=0.0019 at 24-48 hours). There are also significant differences in SAC raw change scores (p=0.0026) and raw scores (p=0.046) for acute concussions and normal performances classified as Possible concussion at 24-48 hours. Among probable concussions at <6 hours and 24-48 hours, acute concussions and normal performances are significantly different in change scores for SCAT symptom severity (p=0.0012-0.0077), SCAT total symptoms (p=0.0093 at <6 hours, p<0.001 at 24-48 hours), and BESS (p=0.0074 at <6 hours, p<0.001 at 24-48 hours). They are also significantly different in SCAT symptom severity raw score at <6 hours ($p<0.001$) and SCAT total symptoms raw score at 24-48 hours ($p<0.001$).

To illustrate how these intraclass differences can be used to inform clinical decision-making, we revisit the examples from the *Classifying Unlikely, Possible, Probable, and Definite Concussion* subsection. Consider the first athlete (19 year-old female) and suppose that her change scores for the SAC, SCAT symptom severity, SCAT total symptoms, and BESS are 0, 6, 4, and -5, respectively. Based on intraclass differences identified in this study for 24-48 hours, there were significant differences between acute concussions and normal performances for the SCAT symptom severity change score, SCAT total symptoms change score, SCAT total symptoms raw score, and the BESS change score. Comparing this athlete's assessments with the mean values for Probable concussions presented in **Table 4** and **Table 5**, we find that the athlete is more comparable to acute concussion in terms of change scores for the SCAT symptom severity and total number of symptoms. Conversely, she is more comparable to the normal performances in terms of the BESS change score. Following the conservative decision-making approaches that are recommended for concussion management, one could treat this athlete as if she has an acute concussion.

Now, consider the 21 year-old male athlete and suppose that his change scores for the SAC and BESS were 0 and 5 respectively. Additionally, his SCAT symptom severity and total symptoms both decreased by 4 compared to baseline. Based on intraclass differences identified in this study for the Possible concussion group at <6 hours, there were significant differences between acute concussions and normal performances in the SCAT symptom severity and SCAT total symptom raw scores. Comparing this athlete's values in these measures (of 1 symptom reported with a severity of 1) to the mean values obtained in our analysis, we find that this athlete more closely resembles the normal performances within the Possible concussions – despite the low SAC raw score and high BESS change score. These results could potentially indicate that additional assessments should be performed on this athlete to confirm the possibility that this athlete is not concussed.

DISCUSSION

Kutcher and Giza proposed a risk-based classification framework for diagnosing acute concussion developed from clinical experience.[4] Compared to traditional binary diagnosis, this framework allows the assessment of acute concussion to reflect the physician's diagnostic certainty. Furthermore, taking this approach allows the injury diagnosis to evolve as the injury evolves and more information becomes available. However, while they provided clinical guidelines for each risk classification, they did not provide specific criteria with respect to commonly recommended and implemented concussion assessment tools. In this research, we designed and evaluated a novel data-driven method for classifying athletes evaluated for acute concussion as either Unlikely, Possible, Probable, and Definite concussion. The major contributions of our research are as follows:

- We develop an objective and data-driven framework which stratifies acute concussion assessment by diagnostic certainty. These risk categories lay the foundation for guiding post-injury management decisions.
- We identify key characteristic which can be used to differentiate between acute concussions and normal performances in each risk category.
- We provide additional, quantitative support for the value of a multidimensional battery, the use of change scores in acute concussion assessment, and the potential implications for several demographic factors and time-of-injury characteristics in acute concussion assessment.

The variables used in our logistic regression and CART models are parts of standard concussion assessment batteries, giving foundation for our framework to be used in sporting environments. To our knowledge, we are the first to combine predictive modeling techniques (i.e., logistic regression and CART) and optimization algorithms to classify athletes into concussion risk categories. Erring in the direction of minimizing false negatives, our framework classified most acute concussions (91.07-91.65%) into the higher risk categories (i.e., Probable and Definite concussion) and most normal performances (73.43-80.77%) into the lower risk categories (i.e., Unlikely and Possible concussion). Additionally, few acute concussions were classified as Unlikely concussion (2.60-3.03%) and few normal performances were classified as Definite concussion (6.15-6.74%).

Our most important finding was that athletes classified as Definite concussion had lower SAC, higher SCAT symptom, and higher BESS scores compared to the other risk categories. In comparing these risk groups, Definite concussions exhibited noticeably more symptoms and greater symptom severity compared to the other risk categories while the Unlikely concussions exhibited mean symptom severities and mean total symptoms close to 0. Definite concussions also had much higher BESS raw scores and lower SAC scores compared to Unlikely concussions. These findings demonstrate the ability of our framework to separate the "easy" cases from the "hard" cases and are consistent with previous research demonstrating symptoms are typically the most sensitive to acute concussion.[8–12] Our findings also provide support for the utility of using neurocognitive assessments and postural control measures for acute concussion assessment, as demonstrated by previous research.[8,26,39–46]

However, among those classified as Possible or Probable concussions, raw scores for SCAT symptom severity and total symptoms are significantly less for acute concussions in the Possible concussion group compared to all baselines ($p<0.01$ for both measures at <6 hours and 24-48 hours using Student's t-test). Additionally, there are no significant differences in the SCAT symptom severity or total symptoms between acute concussions and normal performances in the Probable risk category. These findings demonstrate the difficulty in identifying all acute concussions using symptom raw scores alone. Fortunately, there were some significant differences between acute concussions and normal performances in the Possible and Probable risk categories for change scores in the SAC, SCAT symptom severity, SCAT total symptoms, and the BESS. This result suggests that change scores, which require baseline assessments, have added value when evaluating Possible and Probable concussions and is an important finding regarding the utility of the baseline assessment.

In our analysis, we also sought out to identify differences in athlete demographics and time-of-injury characteristics across and within risk classifications. There were statistically significant differences in age, sex, and number of previous concussions between acute concussions and normal performances within some risk categories. For example, among those classified as Definite concussions, the athletes were, on average, older than those providing normal performances at both <6 hours and 24-48 hour. Outside of age, there were no other consistent demographic differences and risk categories. For time-of-injury variables among acute concussions, a larger proportion of those classified as Unlikely concussion reported the injury immediately and were removed from play immediately compared to those who were classified as Definite concussions. This result suggests that those athletes who were removed from play immediately or assessed immediately after injury may be in the earliest stages of an evolving injury whereby neurocognitive declines, increased symptoms, worsening postural control emerge over time.[16,17,47] However, we note that very few acute concussions were classified as Unlikely concussion, and due to this small sample size, this point may require further investigation.

We also found that baselines comprised most normal performances within the Probable and Definite risk categories. Despite varying parameter settings to balance the sensitivity and specificity of our results, we were unable to drastically improve on the proportion of baselines in these upper risk categories. This finding may be due to performance differences between baseline and unrestricted RTP timepoints. Specifically, the baseline data showed lower SAC scores, higher SCAT symptom assessment scores, and higher BESS scores compared to unrestricted RTP data ($p<0.001$ for all measures and in both training and validation sets). As a result, our logistic regression model categorized the baseline performance of some athletes into the higher risk categories. The performance discrepancy between baseline and unrestricted RTP timepoints is consistent with previous studies[12,16,48–50] and may be attributed to comorbidities[51] or learning effects from multiple assessments prior to return to play.[41,52] Future works may be able to address this shortcoming by incorporating individual items from the SAC, SCAT symptom assessments, and the BESS instead of using total scores. Regardless, this finding highlights the need for clinicians to interpret the administered assessments in the context of the injury, such as an observed mechanism, and differentiate from other injuries and conditions with similar signs and symptoms.[1,4,33,53]

To account for clinical judgment in our methodology, we used a modified range-based computation for sensitivity and specificity, which provided a 6% sensitivity increase and >34% specificity increase in Possible concussion management. Furthermore, the sensitivity of our algorithm mirrors those reported in previous studies (80.0-100.0%) evaluating concussion testing batteries for acute concussion assessment.[8,9,39,54] Methodological differences between our study and these aforementioned studies account for some differences, including the test battery assessments. For example, both Broglio et al.[39] and Resch et al.[9] used the Sensory Organization Test (SOT) for balance assessment instead of the BESS. While both the SOT and BESS reveal similar post-concussion trends in postural control deficits,[40] the SOT has less clinical applicability given its size and cost. Additionally, the diagnosis criteria differed greatly across each study. McCrea et al.,[8] Broglio et al.,[39] and Putukian et al.[54] used different measures of significant change to indicate concussion while Resch et al.[9] used both predictive discriminant analyses and clinical interpretation guidelines. In comparison, we paired a data-driven optimization framework with predictive modeling methods (i.e., logistic regression and CART) to classify athletes into risk categories. By using predictive modeling methods, we were able to simultaneously incorporate demographic information and time-of-injury characteristics, along with SAC, SCAT symptoms, and BESS results. Finally, the concussed sample used in the present study (n=1085 for <6 hours and n=1413 for 24-48 hours) is much larger than those in the aforementioned studies (n=32-166).

From a clinical perspective, previous studies have discussed the importance and value of taking a heterogeneous and targeted approach to concussion management.[55–57] However, since the focus of this study was on identifying acute concussion, it does not address injury heterogeneity by accounting for potential concussion subtypes or clinical profiles. However, our work lays the foundation to do so using clustering or clinically determined approaches.

Our study is not without limitations. First, we acknowledge that our framework does not provide a recommendation for post-injury management for athletes classified in each risk category. These post-injury decisions are beyond the scope of our study and are an important topic for future research. To this end, clinicians can still benefit from knowing the degree of certainty in a diagnosis decision before determining the next course of action. Second, our study treats all concussions in the CARE data as truly concussed, regardless of the medical staff certainty. Thus, there is the possibility that our models were trained and validated on athletes who were not actually concussed but were labeled so. Third, the differences between our training and validation data in important clinical measures such as PTA or LOC may have caused differences in the presentation of concussion between those two groups, potentially explaining some of the prediction errors in our models. Determining training and validation data using random subset selection instead of by a time-based cut-off could lead to a more homogeneous division in data and ultimately, improved modeling results. Fourth, since our study data only included athletes aged 18-22, we cannot directly apply our results to populations beyond this group. Therefore, future studies should focus on other population groups, such as youth sports and professional athletes, to determine the generalizability of our results beyond our study population. Fifth, we were limited in our ability to include change scores and time-of-injury characteristics in our models, as these measures were not available for baseline data. Finding ways to incorporate such variables in future analysis may improve our results. Furthermore, our analysis focused on the SAC, SCAT symptom assessments, and the BESS. Data limitations precluded our ability to include assessments such as the Sensory Organization Test, computer based neurocognitive testing, the King-Devick test, and/or the Vestibular/Ocular Motor Screening Assessment that have shown promise in other investigations,[58–61]. Finally, as there is no gold standard for concussion diagnosis, we did not have a comparative mechanism for our results.

The objective, algorithmic approach we proposed and developed for risk-based classification of athletes undergoing acute concussion assessment extends the original framework proposed by Kutcher and Giza.[4] By applying predictive modeling and optimization methods, our work provides a promising first-step in taking an evidence-based approach to acute concussion assessment stratification. While the clinical examination remains the gold standard for concussion diagnosis, the models we have designed and analyzed have the potential to provide valuable decision support for clinicians.

ACKNOWLEDGMENTS

The material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE 1256260. This publication was made possible, in part, with support from the Grand Alliance Concussion Assessment, Research, and Education (CARE) Consortium, funded, in part, by the National College Athletic Association (NCAA) and the Department of Defense (DOD). The U.S. Army Medical Research Acquisition Activity, 820 Chandler Street, Fort Detrick MD 21702-5014 is the awarding and administering acquisition office. This work was supported by the Office of the Assistant Secretary of Defense for Health Affairs through the Psychological Health and Traumatic Brain Injury Program under Award No. W81XWH-14-2-0151. Opinions, interpretations, conclusions and recommendations are those of the author(s) and are not necessarily endorsed by the Department of Defense (DHP funds).

April Marie (Reed) Hoy, MS, ATC (Azusa Pacific University), Joseph B. Hazzard Jr., EdD, ATC (Bloomsburg University), Louise A. Kelly, PhD (California Lutheran University), Justus D. Ortega, PhD (Humboldt State University), Nicholas Port, PhD (Indiana University), Margot Putukian, MD (Princeton University), Gerald McGinty, DPT and Jonathan C. Jackson, PhD (United States Air Force Academy), Kenneth L. Cameron, PhD, MPC, ATC (United States Military Academy), Christopher Giza, MD (University of California Los Angeles), Holly J. Benjamin, MD (University of Chicago), Thomas Buckley, EdD, ATC and Thomas W. Kaminski, PhD, ATC (University of Delaware), James R. Clugston, MD, MS (University of Florida), Julianne D. Schmidt, PhD, ATC (University of Georgia), Louis A. Feigenbaum, DPT, ATC (University of Miami), James T. Eckner, MD, MS (University of Michigan), Kevin M. Guskiewicz, PhD, ATC and Jason P. Mihalik, PhD, ATC (University of North Carolina), Jessica Dysart Miles, PhD, ATC (University of North Georgia), Scott Anderson, ATC (University of Oklahoma), Christina L. Master, MD (University of Pennsylvania), Anthony P. Kontos, PhD and Micky Collins, PhD (University of Pittsburgh), Sara P.D. Chrisman, MD, MPH (University of Washington), Alison Brooks, MD, MPH (University of Wisconsin), Steven Rowson, PhD (Virginia Tech), Christopher M. Miles, MD and Laura J. Lintner, DO (Wake Forest University), Brian H. Dykhuizen, MS, ATC, LAT (Wilmington College).

AUTHOR DISCLOSURE STATEMENT

No competing financial interests exist.

REFERENCES

1. McCrory, P., Meeuwisse, W., Dvorak, J., Aubry, M., Bailes, J., Broglio, S., Cantu, R.C., Cassidy, D., Echemendia, R.J., Castellani, R.J., and others. (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. , bjsports-2017-097699.

2. Meehan, S.K., Mirdamadi, J.L., Martini, D.N., and Broglio, S.P. (2017). Changes in Cortical Plasticity in Relation to a History of Concussion during Adolescence. Front. Hum. Neurosci. 11, 1–13.

3. Makdissi, M., Davis, G., and McCrory, P. (2015). Clinical challenges in the diagnosis and assessment of sports-related concussion. Neurol. Clin. Pract. 5, 2–5.

4. Kutcher, J.S., and Giza, C.C. (2014). Sports Concussion Diagnosis and Management. Continuum (N. Y). 20, 1552–1569.

5. McCrory, P., Meeuwisse, W.H., Aubry, M., Cantu, R.C., Dvorák, J., Echemendia, R.J., Engebretsen, L., Johnston, K.M., Kutcher, J.S., Raftery, M., Sills, A., Benson, B.W., Davis, G. a., Ellenbogen, R., Guskiewicz, K.M., Herring, S. a., Iverson, G.L., Jordan, B.D., Kissick, J., McCrea, M., McIntosh, A.S., Maddocks, D.L., 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, 255–279.

6. Broglio, S.P., Cantu, R.C., Gioia, G.A., Guskiewicz, K.M., Kutcher, J., Palm, M., and McLeod, T.C.V. (2014). National Athletic Trainers' Association Position Statement: Management of Sport Concussion. J. Athl. Train. 49, 245–265.

7. Harmon, K.G., Drezner, J.A., Gammons, M., Guskiewicz, K.M., Halstead, M., Herring, S.A., Kutcher,

J.S., Pana, A., Putukian, M., and Roberts, W.O. (2013). American Medical Society for Sports Medicine position statement: concussion in sport. Br. J. Sports Med. 47, 15–26.

8. McCrea, M.A., Barr, W.B., Guskiewicz, K.M., Randolph, C., Marshall, S.W., Cantu, R.C., Onate, J.A., and Kelly, J.P. (2005). Standard regression-based methods for measuring recovery after sport-related concussion. J. Int. Neuropsychol. Soc. 11, 58–69.

9. Resch, J.E., Brown, C.N., Schmidt, J., Macciocchi, S.N., Blueitt, D., Cullum, C.M., and Ferrara, M.S. (2016). The sensitivity and specificity of clinical measures of sport concussion: three tests are better than one. BMJ Open Sport Exerc. Med. 2, e000012.

10. Register-Mihalik, J.K., Guskiewicz, K.M., Mihalik, J.P., Schmidt, J.D., Kerr, Z.Y., and McCrea, M. a. (2013). Reliable Change, Sensitivity, and Specificity of a Multidimensional Concussion Assessment Battery. J. Head Trauma Rehabil. 28, 274–283.

11. Chin, E.Y., Nelson, L.D., Barr, W.B., McCrory, P., and McCrea, M.A. (2016). Reliability and validity of the sport concussion assessment tool-3 (SCAT3) in high school and collegiate athletes. Am. J. Sports Med. 44, 2276–2285.

12. Garcia, G.-G.P., Broglio, S.P., Lavieri, M.S., McCrea, M., and McAllister, T. (2018). Quantifying the Value of Multidimensional Assessment Models for Acute Concussion: An Analysis of Data from the NCAA-DoD Care Consortium. Sport. Med. 48, 1739–1749.

13. Williamson, I.J.S., and Goodman, D. (2006). Converging evidence for the under-reporting of concussions in youth ice hockey. Br. J. Sports Med. 40, 128–132.

14. McCrea, M., Hammeke, T., Olsen, G., Leo, P., and Guskiewicz, K. (2004). Unreported concussion in high school football players: implications for prevention. Clin J Sport Med 14, 13–17.

15. Llewellyn, T., Burdette, G.T., Joyner, A.B., and Buckley, T.A. (2014). Concussion reporting rates at

the conclusion of an intercollegiate athletic career. Clin. J. Sport Med. 24, 76–79.

16. McCrea, M., Guskiewicz, K.M., Marshall, S.W., Barr, W., Randolph, C., Cantu, R.C., Onate, J. a, Kelly, J.P., and Page, P. (2004). Acute effects and recovery time following concussions in collegiate football players. JAMA 290, 2556–2563.

17. Makdissi, M., Darby, D., Maruff, P., Ugoni, A., Brukner, P., and McCrory, P.R. (2010). Natural History of Concussion in Sport. Am. J. Sports Med. 38, 464–471.

18. McDonald, W.I., Compston, A., Edan, G., Goodkin, D., Hartung, H.-P., Lublin, F.D., McFarland, H.F., Paty, D.W., Polman, C.H., Reingold, S.C., Sandberg-Wollheim, M., Sibley, W., Thompson, A., Van Den Noort, S., Weinshenker, B.Y., and Wolinsky, J.S. (2001). Recommended diagnostic criteria for multiple sclerosis: Guidelines from the international panel on the diagnosis of multiple sclerosis. Ann. Neurol. 50, 121–127.

19. McKhann, G.M., Knopman, D.S., Chertkow, H., Hyman, B.T., Jack, C.R., Kawas, C.H., Klunk, W.E., Koroshetz, W.J., Manly, J.J., Mayeux, R., and others. (2011). The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's Dement. 7, 263–269.

20. American Diabetes Society. (2016). 2. Classification and Diagnosis of Diabetes. Diabetes Care 39, S13–S22.

21. Broglio, S.P., McCrea, M., McAllister, T., Harezlak, J., Katz, B., Hack, D., and Hainline, B. (2017). A National Study on the Effects of Concussion in Collegiate Athletes and US Military Service Academy Members: The NCAA–DoD Concussion Assessment, Research and Education (CARE) Consortium Structure and Methods. Sport. Med. 47, 1437–1451.

22. Carney, N., Ghajar, J., Jagoda, A., Bedrick, S., Davis-O'Reilly, C., Du Coudray, H., Hack, D., Helfand, N., Huddleston, A., Nettleton, T., and Riggio, S. (2014). Concussion guidelines step 1: Systematic review of prevalent indicators. Neurosurgery 75, 3–15.

23. Royston, P. (2004). Multiple imputation of missing values. Stata J. 4, 224–241.

24. Covassin, T., Buz Swanik, C., and Sachs, M.L. (2003). Sex Differences and the Incidence of Concussions Among Collegiate Athletes. J. Athl. Train. 38, 238–244.

25. Broshek, D.K., Kaushik, T., Freeman, J.R., Erlanger, D., Webbe, F., and Barth, J.T. (2005). Sex differences in outcome following sports-related concussion. J. Neurosurg. 102, 856–863.

26. Covassin, T., Schatz, P., and Swanik, C.B. (2007). Sex differences in neuropsychological function and post-concussion symptoms of concussed collegiate athletes. Neurosurgery 61, 345–350.

27. Covassin, T., Elbin, R.J., Harris, W., Parker, T., and Kontos, A. (2012). The Role of Age and Sex in Symptoms, Neurocognitive Performance, and Postural Stability in Athletes After Concussion. Am. J. Sports Med. 40, 1303–1312.

28. Kutcher, J.S., and Eckner, J.T. (2010). At-risk populations in sports-related concussion. Curr. Sports Med. Rep. 9, 16–20.

29. Gessel, L.M., Fields, S.K., Collins, C.L., Dick, R.W., and Comstock, R.D. (2007). Concussions among United States high school and collegiate athletes. J. Athl. Train. (National Athl. Trainers' Assoc. 42, 495–503.

30. Asken, B.M., McCrea, M.A., Clugston, J.R., Snyder, A.R., Houck, Z.M., and Bauer, R.M. (2016). "Playing Through It": Delayed Reporting and Removal From Athletic Activity After Concussion Predicts Prolonged Recovery. J. Athl. Train. 51, 329–335.

31. Elbin, R.J., Sufrinko, A.M., Schatz, P., French, J., Henry, L., Burkhart, S., Collins, M.W., and Kontos, A.P. (2016). Removal From Play After Concussion and Recovery Time. Pediatrics 138, e20160910–e20160910.

32. Echemendia, R.J., Bruce, J.M., Meeuwisse, W., Hutchison, M.G., Comper, P., and Aubry, M. (2017). Can visible signs predict concussion diagnosis in the National Hockey League? Br. J. Sports Med. , bjsports-2016-097090.

33. Bruce, J.M., Echemendia, R.J., Meeuwisse, W., Hutchison, M.G., Aubry, M., and Comper, P. (2017). Development of a risk prediction model among professional hockey players with visible signs of concussion. Br. J. Sports Med. , bjsports-2016-097091.

34. McCrea, M., Kelly, J.P., Randolph, C., Kluge, J., Bartolic, E., Finn, G., and Baxter, B. (1998). Standardized assessment of concussion (SAC): on-site mental status evaluation of the athlete. J. Head Trauma Rehabil. 13, 27–35.

35. Concussion in Sport Group. (2013). Sport Concussion Assessment Tool - 3rd Edition. Br. J. Sports Med. 47, 259.

36. Riemann, B.L., Guskiewicz, K.M., and Shields, E.W. (1999). Relationship between Clinical and Forceplate Measures of Postural Stability. J. Sport Rehabil. 8, 71–82.

37. Breiman, L., Friedman, J., Stone, C.J., and Olshen, R.A. (1984). Classification and regression trees. CRC press.

38. He, H., Bai, Y., Garcia, E.A., and Li, S. (2008). ADASYN: Adaptive synthetic sampling approach for imbalanced learning. Proc. Int. Jt. Conf. Neural Networks , 1322–1328.

39. Broglio, S.P., Macciocchi, S.N., and Ferrara, M.S. (2007). Sensitivity of the concussion assessment battery. Neurosurgery 60, 1050–1057.

40. Guskiewicz, K.M., Ross, S.E., and Marshall, S.W. (2001). Postural Stability and Neuropsychological Deficits After Concussion in Collegiate Athletes. J. Athl. Train. 36, 263–273.

41. Valovich McLeod, T.C., Perrin, D.H., Guskiewicz, K.M., Shultz, S.J., Diamond, R., and Gansneder, B.M. (2004). Serial administration of clinical concussion assessments and learning effects in healthy young athletes. Clin. J. Sport Med. 14, 287–295.

42. Guskiewicz, K.M. (2001). Postural Stability Assessment Following Concussion: One Piece of the Puzzle. Clin. J. Sport Med. 11, 182–189.

43. Riemann, B.L., and Guskiewicz, K.M. (2000). Effects of mild head injury on postural stability as measured through clinical balance testing. J. Athl. Train. 35, 19–25.

44. Buckley, T.A., Munkasy, B.A., and Clouse, B.P. (2017). Sensitivity and Specificity of the Modified Balance Error Scoring System in Concussed Collegiate Student Athletes. Clin. J. Sport Med. , 1.

45. Barr, W.B., and McCrea, M. (2001). Sensitivity and specificity of standardized neurocognitive testing immediately following sports concussion. J. Int. Neuropsychol. Soc. 7, 693–702.

46. Sufrinko, A., McAllister-Deitrick, J., Womble, M., and Kontos, A. (2017). Do Sideline Concussion Assessments Predict Subsequent Neurocognitive Impairment After Sport-Related Concussion? J. Athl. Train. 52, 1062–6050–52.4.01.

47. Guskiewicz, K.M., McCrea, M., Marshall, S.W., Cantu, R.C., Randolph, C., Barr, W., Onate, J.A., and Kelly, J.P. (2003). Cumulative Effects Associated With Recurrent Concussion in Collegiate Football Players. J. Am. Med. Assoc. 290, 2549.

48. McCrea, M., Guskiewicz, K., Randolph, C., Barr, W.B., Hammeke, T. a., Marshall, S.W., Powell, M.R., Woo Ahn, K., Wang, Y., and Kelly, J.P. (2013). Incidence, Clinical Course, and Predictors of Prolonged Recovery Time Following Sport-Related Concussion in High School and College

Athletes. J. Int. Neuropsychol. Soc. 19, 22–33.

49. Shehata, N., Wiley, J.P., Richea, S., Benson, B.W., Duits, L., and Meeuwisse, W.H. (2009). Sport concussion assessment tool: Baseline values for varsity collision sport athletes. Br. J. Sports Med. 43, 730–734.

50. Piland, S.G., Ferrara, M.S., Macciocchi, S.N., Broglio, S.P., and Gould, T.E. (2010). Investigation of baseline self-report concussion symptom scores. J. Athl. Train. 45, 273–278.

51. Lovell, M.R., Iverson, G.L., Collins, M.W., Podell, K., Johnston, K.M., Pardini, D., Pardini, J., Norwig, J., and Maroon, J.C. (2006). Measurement of Symptoms Following Sports-Related Concussion: Reliability and Normative Data for the Post-Concussion Scale. Appl. Neuropsychol. 13, 166–174.

52. Moreau, M.S., Langdon, J., and Buckley, T. a. (2014). The lived experience of an in-season concussion amongst NCAA Division I student-athletes. Int. J. Exerc. Sci. 7, 62–74.

53. Zuckerman, S.L., Totten, D.J., Rubel, K.E., Kuhn, A.W., Yengo-Kahn, A.M., and Solomon, G.S. (2016). Mechanisms of Injury as a Diagnostic Predictor of Sport-Related Concussion Severity in Football, Basketball, and Soccer. Neurosurgery 63, 102–112.

54. Putukian, M., Echemendia, R., Dettwiler-Danspeckgruber, A., Duliba, T., Bruce, J., Furtado, J.L., and Murugavel, M. (2015). Prospective clinical assessment using sideline concussion assessment tool-2 testing in the evaluation of sport-related concussion in college athletes. Clin. J. Sport Med. 25, 36–42.

55. Collins, M.W., Kontos, A.P., Reynolds, E., Murawski, C.D., and Fu, F.H. (2014). A comprehensive, targeted approach to the clinical care of athletes following sport-related concussion. Knee Surgery, Sport. Traumatol. Arthrosc. 22, 235–246.

56. Collins, M.W., Kontos, A.P., Okonkwo, D.O., Almquist, J., Bailes, J., Barisa, M., Bazarian, J., Bloom,

O.J., Brody, D.L., Cantu, R., Cardenas, J., Clugston, J., Cohen, R., Echemendia, R., Elbin, R.J., Ellenbogen, R., Fonseca, J., Gioia, G., Guskiewicz, K., Heyer, R., Hotz, G., Iverson, G.L., Jordan, B., Manley, G., Maroon, J., McAllister, T., McCrea, M., Mucha, A., Pieroth, E., Podell, K., Pombo, M., Shetty, T., Sills, A., Solomon, G., Thomas, D.G., Valovich McLeod, T.C., Yates, T., and Zafonte, R. (2016). Statements of Agreement From the Targeted Evaluation and Active Management (TEAM) Approaches to Treating Concussion Meeting Held in Pittsburgh, October 15-16, 2015. Neurosurgery 79, 912–929.

57. Ellis, M.J., Ryner, L.N., Sobczyk, O., Fierstra, J., Mikulis, D.J., Fisher, J.A., Duffin, J., and Mutch, W.A.C. (2016). Neuroimaging Assessment of Cerebrovascular Reactivity in Concussion: Current Concepts, Methodological Considerations, and Review of the Literature. Front. Neurol. 7, 1–16.

58. Mucha, A., Collins, M.W., Elbin, R.J., Furman, J.M., Troutman-Enseki, C., DeWolf, R.M., Marchetti, G., and Kontos, A.P. (2014). A Brief Vestibular/Ocular Motor Screening (VOMS) Assessment to Evaluate Concussions. Am. J. Sports Med. 42, 2479–2486.

59. Pearce, K.L., Sufrinko, A., Lau, B.C., Henry, L., Collins, M.W., and Kontos, A.P. (2015). Near Point of Convergence After a Sport-Related Concussion. Am. J. Sports Med. 43, 3055–3061.

60. Kontos, A.P., Sufrinko, A., Elbin, R.J., Puskar, A., and Collins, M.W. (2016). Reliability and Associated Risk Factors for Performance on the Vestibular/Ocular Motor Screening (VOMS) Tool in Healthy Collegiate Athletes. Am. J. Sports Med. 44, 1400–1406.

61. Anzalone, A.J., Blueitt, D., Case, T., McGuffin, T., Pollard, K., Garrison, J.C., Jones, M.T., Pavur, R., Turner, S., and Oliver, J.M. (2017). A Positive Vestibular/Ocular Motor Screening (VOMS) Is Associated With Increased Recovery Time After Sports-Related Concussion in Youth and Adolescent Athletes. Am. J. Sports Med. 45, 474–479.

Table 1. Data characteristics of training and validation set with respect to each timepoint

	Baseline		<6 hours		24-48 hours		Unrestricted RTP	
	Training	*Validation*	*Training*	*Validation*	*Training*	*Validation*	*Training*	*Validation*
n	751	884	546	539	719	694	716	629
Height in meters (SD)	1.79 (0.11)**	1.78 (0.12)	1.80 (0.12)**	1.79 (0.11)	1.79 (0.12)*	1.78 (0.12)	1.79 (0.12)*	1.78 (0.12)
Weight in kg (SD)	83.75 (21.65)**	81.95 (21.25)	86.01 (22.45)**	83.32 (21.39)	83.94 (22.08)**	81.33 (20.85)	83.33 (21.4)**	80.74 (21.01)
Age in years (SD)	19.43 (1.30)*	19.24 (1.32)	19.36 (1.32)	19.23 (1.36)	19.34 (1.27)**	19.2 (1.33)	19.37 (1.30)	19.27 (1.32)
Male Sex (% yes)	60.80%	57.58%	64.47%	61.6%	60.08%	56.34%	60.06%**	54.05%
Number of previous concussions (SD)	0.73 (1.0)*	0.59 (0.83)	0.78 (1.03)*	0.57 (0.78)	0.74 (1.03)*	0.58 (0.82)	0.70 (0.93)*	0.52 (0.77)
Report injury immediately? (% yes)[1]	NA	NA	55.31%**	60.67%	40.47%**	46.54%	41.2%**	47.22%
Removed from play immediately? (% yes)[1]	NA	NA	57.51%	56.77%	46.18%	47.26%	48.6%	45.31%
LOC? (% yes)[1]	NA	NA	5.86%	5.01%	4.45%	4.76%	5.45%	4.77%
PTA? (% yes)[1]	NA	NA	12.27%**	9.09%	11.4%**	8.65%	11.45%*	7.63%
RGA? (% yes)[1]	NA	NA	5.86%	5.57%	5.84%	5.33%	5.59%	4.13%
SAC change score (SD)[1]	NA	NA	-0.82 (3.19)*	-1.76 (3.58)	-0.4 (2.61)*	-1.11 (2.72)	0.97 (2.13)*	0.51 (1.99)
SAC raw score (SD)	27.05 (2.01)*	27.38 (1.96)	26.18 (2.94)**	25.83 (3.33)	26.63 (2.41)	26.48 (2.77)	27.93 (1.75)*	28.19 (1.76)
SCAT symptom severity change score (SD)[1]	NA	NA	23.22 (20.75)	23.45 (22.76)	19.48 (21.68)	20.62 (23.27)	-4.87 (8.8)	-5.12 (9.71)
SCAT symptom severity raw score (SD)	5.16 (8.54)	5.25 (9.71)	28.64 (20.83)	28.58 (21.52)	25.21 (21.54)	26.04 (21.93)	0.63 (1.97)	0.46 (1.67)
SCAT total symptoms change score (SD)[1]	NA	NA	8.01 (5.93)	8.03 (6.43)	7.49 (6.58)	8.05 (6.88)	-2.48 (4)	-2.55 (3.83)
SCAT total symptoms raw score (SD)	2.81 (3.85)	2.77 (3.89)	10.86 (5.42)	10.77 (5.57)	10.53 (6.03)	10.91 (6.18)	0.47 (1.39)**	0.34 (1.09)
BESS change score (SD)[1]			3.58 (8.64)	4.02 (8.52)	1.5 (7.48)**	2.58 (8.19)	-2.35 (6.32)	-2.39 (5.9)
BESS raw score (SD)	12.7 (6.32)	12.78 (6.21)	16.35 (8.78)**	17.36 (8.38)	14.42 (7.87)*	15.82 (8.1)	10.46 (5.81)	10.84 (5.36)

Change score at a time point is computed as: raw score at timepoint - raw score at baseline

*Significantly different from validation data at same time point based on Student's t-test ($p<0.01$)

** Significantly different from validation data at same time point based on Student's t-test ($p<0.05$)

n, number of data points; SD, standard deviation; SAC, Standard Assessment of Concussion; SCAT, Sport Concussion Assessment Tool; BESS, Balance Error Scoring System; NA, variable not available for baseline data

Table 2. Multivariate logistic regression coefficients at <6 hours and 24-48 hours

	<6 hours			24-48 hours		
Study Variable	*Coefficient*	*SE*	*p-value*	*Coefficient*	*SE*	*p-value*
Intercept	-0.57	1.71	0.74	-0.37	1.46	0.80
Male Sex	0.71	0.27	0.01	0.45	0.22	0.04
Report injury immediately?	-0.46	0.33	0.16	-1.37	0.22	<0.01
Removed from play immediately?	-0.91	0.35	0.01	NA	NA	NA
SAC raw score	-0.09	0.06	0.13	-0.06	0.05	0.23
SCAT symptom severity raw score	0.06	0.03	0.02	-0.05	0.02	0.02
SCAT total symptoms raw score	0.28	0.07	<0.01	0.47	0.05	<0.01
BESS raw score	0.03	0.02	0.08	0.00	0.02	0.94

SE, standard error; SAC, Standard Assessment of Concussion; SCAT, Sport Concussion Assessment Tool; BESS, Balance Error Scoring System; NA, variable not included in this model

Table 3. Distribution of acute concussions and normal performances among risk classifications at <6 hours and 24-48 hours

		Unlikely		Possible		Probable		Definite	
		n	*%*	*n*	*%*	*n*	*%*	*n*	*%*
<6 hours	*Acute concussion**	14	2.60%	31	5.75%	60	11.13%	434	80.52%
	Normal performance	696	46.00%	526	34.77%	189	12.49%	102	6.74%
	*(Unrestricted RTP)***	282	44.83%	329	52.31%	15	2.38%	3	0.48%
	(Baseline)	414	46.83%	197	22.29%	174	19.68%	99	11.20%
	Total	710	34.60%	557	27.14%	249	12.13%	536	26.12%
24-48 hours	*Acute concussion**	21	3.03%	41	5.91%	110	15.85%	522	75.22%
	Normal performance	714	47.19%	397	26.24%	309	20.42%	93	6.15%
	*(Unrestricted RTP)***	275	43.72%	312	49.60%	36	5.72%	6	0.95%
	(Baseline)	439	49.66%	85	9.62%	273	30.88%	87	9.84%
	Total	735	33.30%	438	19.85%	419	18.99%	615	27.87%

*Distributions of acute concussions and normal performances within risk classifications are significantly different at p<0.001 using Kolmogorov-Smirnov test

**Distributions of Unrestricted RTP and Baseline timepoints within risk classifications are significantly different at p<0.01 using Kolmogorov-Smirnov test

Table 4. Comparison of study variables for acute concussions classified as *Unlikely*, *Possible*, *Probable,* and *Definite* concussion at <6 hours and 24-48 hours

	<6 hours				**24-48 hours**			
	Unlikely	*Possible*	*Probable*	*Definite*	*Unlikely*	*Possible*	*Probable*	*Definite*
n	14	31	60	434	21	41	110	522
Age in years (SD)	19.93 (0.80)*	19.42 (1.48)	19.23 (1.28)	19.19 (1.37)*	19.38 (1.29)	19.20 (1.40)	19.15 (1.29)	19.20 (1.33)*
Male Sex (% yes)	78.57%*	74.19%	43.33%	62.67%	61.90%	73.17%*[2]	49.09%	56.32%
Number of previous concussions (SD)	0.43 (0.62)	0.55 (0.87)	0.60 (0.82)	0.56 (0.77)	0.52 (0.73)	0.66 (0.95)	0.53 (0.72)	0.58 (0.82)*
Report injury immediately? (% yes)	100.00%*[1]	67.74%*[1]	83.33%*	55.76%	100.00%*[1]	51.22%*[1,2]	76.36%*	37.74%
Removed from play immediately? (% yes)	100.00%*[1]	64.52%*[1]	76.67%*	52.07%	90.48%*[1]	51.22%*[1]	62.73%*	41.95%
LOC? (% yes)	14.29%	6.45%*	5.00%	4.61%*	9.52%	4.88%*	3.64%	4.79%*
PTA? (% yes)	7.14%	3.23%*	3.33%	10.37%*	9.52%	7.32%	5.45%	9.39%
RGA? (% yes)	21.43%	6.45%*	0.00%	5.76%*	9.52%*	2.44%	7.27%	4.98%*
SAC change score (SD)	0.14 (1.46)	-0.97 (2.86)[1]	-0.55 (2.10)	-1.76 (3.51)	-0.52 (2.32)	-1.59 (2.56)*[1,2]	-0.15 (1.63)	-1.28 (2.76)*
SAC raw score (SD)	28.57 (1.18)[1]	26.61 (2.88)[1]	27.48 (1.80)	25.49 (3.42)*	27.29 (1.72)	25.66 (2.69)*[1,2]	27.76 (1.33)	26.24 (2.90)*
SCAT symptom severity change score (SD)	-2.79 (5.83)[1]	-0.48 (9.08)[1]	5.18 (7.17)*[1]	28.49 (21.94)*	-3.81 (7.40)[1]	-1.00 (10.92)[1]	4.38 (8.40)*[1]	26.42 (23.33)*
SCAT symptom severity raw score (SD)	0.71 (1.33)[1]	2.74 (1.24)*[1]	8.70 (3.00)*[1]	34.01 (20.29)*	0.90 (1.72)*[1]	3.24 (4.11)*[1]	7.15 (5.47)[1]	32.88 (20.95)*
SCAT total symptoms change score (SD)	-1.21 (3.41)[1]	0.42 (3.98)[1]	2.60 (3.51)*[1]	9.63 (5.85)*	-2.05 (2.87)[1,2]	-0.54 (3.79)[1,2]	2.63 (3.47)*[1]	10.21 (6.14)*
SCAT total symptoms raw score (SD)	0.57 (0.90)[1,2]	2.06 (1.08)*[1,2]	4.73 (1.42)[1]	12.57 (4.53)*	0.33 (0.56)*[1,2]	1.59 (1.56)*[1,2]	4.28 (1.79)*[1]	13.47 (4.74)*
BESS change score (SD)	-1.93 (8.92)[1]	3.97 (7.45)	2.87 (6.39)*	4.94 (9.41)	0.62 (5.60)	0.83 (5.92)	1.46 (6.73)*	3.52 (8.40)*
BESS raw score (SD)	11.64 (6.00)[1]	15.00 (6.51)[1]	13.83 (7.34)	18.40 (9.02)*	11.62 (5.62)[1]	13.17 (6.29)[1]	14.07 (7.18)[1]	16.89 (8.61)

Change score at a timepoint is computed as: raw score at timepoint - raw score at baseline

*Significantly different ($p<0.05$) from normal performances in the same risk classification and timepoint based on Student's t-test

[1]Significantly different ($p<0.05$) from *Definite* concussion at the same timepoint based on Tukey's post-hoc pairwise comparisons

[2]Significantly different ($p<0.05$) from *Probable* concussion at the same timepoint based on Tukey's post-hoc pairwise comparisons

n, number of data points; SD, standard deviation; LOC, loss of consciousness; PTA, post-traumatic amnesia; RGA, retrograde amnesia; SAC, Standard Assessment of Concussion; SCAT, Sport Concussion Assessment Tool; BESS, Balance Error Scoring System

Table 5. Comparison of study variables for normal performances classified as *Unlikely*, *Possible*, *Probable,* and *Definite* concussion at <6 hours and 24-48 hours

	<6 hours				**24-48 hours**			
	Unlikely	*Possible*	*Probable*	*Definite*	*Unlikely*	*Possible*	*Probable*	*Definite*
n (% baseline)	696 (59.48%)	526 (37.45%)	189 (92.06%)	102 (97.06%)	714 (61.48%)	397 (21.41%)	309 (88.35%)	93 (93.55%)
Age in years (SD)	19.28 (1.27)*	19.33 (1.37)	19.13 (1.32)	18.86 (1.30)*[1]	19.36 (1.31)	19.16 (1.30)	19.24 (1.35)	18.86 (1.29)*[1]
Male Sex (% yes)	48.13%*	72.24%[1]	41.80%[2]	53.92%	56.16%	59.19%*	54.05%	49.46%
Number of previous concussions (SD)	0.55 (0.81)	0.53 (0.76)	0.58 (0.86)	0.70 (0.84)	0.54 (0.78)	0.51 (0.80)	0.61 (0.82)	0.76 (0.88)*
Report injury immediately? (% yes)**	84.75%*	14.89%*	53.33%*	33.33%	100.00%*	2.88%*	30.56%*	33.33%
Removed from play immediately? (% yes)**	84.40%*	11.85%*	46.67%*	33.33%	77.09%*	18.59%*	33.33%*	50.00%
LOC? (% yes)**	7.45%	2.43%*	6.67%	0.00%*	8.73%	0.96%*	8.33%	0.00%*
PTA? (% yes)**	7.09%	7.90%*	13.33%	0.00%*	8.73%	6.41%	8.33%	16.67%
RGA? (% yes)**	6.03%	2.43%*	6.67%	0.00%*	5.09%*	3.21%	5.56%	0.00%*
SAC change score (SD)	0.85 (1.93)	0.54 (2.13)	0.13 (1.82)	-0.33 (1.89)	0.85 (2.03)	0.54 (2.07)*	0.28 (1.74)	1.17 (2.19)*
SAC raw score (SD)	28.11 (1.60)	27.35 (2.25)[1]	27.40 (1.90)[1]	27.27 (1.93)[1]	27.88 (1.82)	27.30 (2.39)*[1]	27.92 (1.40)[2]	27.34 (2.02)*
SCAT symptom severity change score (SD)	-4.98 (8.37)	-5.13 (10.12)	-8.33 (18.17)*	-8.67 (14.27)*	-5.09 (8.96)	-4.99 (8.71)	-5.75 (17.90)*	-12.83 (16.08)*
SCAT symptom severity raw score (SD)	0.41 (1.02)	1.02 (1.41)*	8.22 (3.23)[1,2]	25.65 (15.51)*[1,2]	0.34 (0.96)*	0.99 (2.29)*	6.64 (5.50)[1,2]	24.75 (16.6)*[1,2]
SCAT total symptoms change score (SD)	-2.63 (3.51)	-2.54 (3.90)	-1.73 (5.96)*	-1.67 (4.64)*	-2.61 (3.64)	-2.68 (3.73)	-1.06 (4.52)*	-2.50 (6.85)*
SCAT total symptoms raw score (SD)	0.28 (0.64)	0.81 (1.16)*[1]	4.84 (1.59)[1,2]	11.41 (3.99)*[1,2]	0.20 (0.49)*	0.66 (1.37)*[1]	3.85 (1.82)*[1,2]	11.87 (3.85)*[1,2]
BESS change score (SD)	-2.72 (6.03)	-1.64 (5.68)	-1.60 (5.55)*	-0.33 (3.68)*	-2.06 (6.19)	-2.08 (5.50)	-3.06 (6.42)*	-0.83 (3.02)*
BESS raw score (SD)	10.74 (5.14)	12.96 (5.97)[1]	13.20 (7.05)[1]	16.41 (7.77)[1,2]	11.76 (5.68)	11.28 (5.46)	13.38 (6.76)[1,2]	15.56 (7.67)*[1,2]

Change score at a timepoint is computed as: raw score at timepoint - raw score at baseline

*Significantly different ($p<0.05$) from acute concussions in the same risk classification and timepoint based on Student's t-test

**Variable not available for baseline data

[1]Significantly different ($p<0.05$) from *No* concussion at the same timepoint based on Tukey's post-hoc pairwise comparisons

[2]Significantly different ($p<0.05$) from *Possible* concussion at the same timepoint based on Tukey's post-hoc pairwise comparisons

n, number of data points; SD, standard deviation; LOC, loss of consciousness; PTA, post-traumatic amnesia; RGA, retrograde amnesia; SAC, Standard Assessment of Concussion; SCAT, Sport Concussion Assessment Tool; BESS, Balance Error Scoring System