Vision Research 200 (2022) 108081

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Optimizing VOMS for identifying acute concussion in collegiate athletes: Findings from the NCAA-DoD CARE consortium

Lyndsey M. Ferris [a], Anthony P. Kontos [b], Shawn R. Eagle [b], R.J. Elbin [c], James R. Clugston [d], Justus Ortega [e], Nicholas L. Port [a,*]

[a] Indiana University School of Optometry, United States
[b] University of Pittsburgh, United States
[c] University of Arkansas, United States
[d] University of Florida, United States
[e] Humboldt State University Kinesiology and Recreation Administration, United States

ARTICLE INFO

Keywords:
Sport-related concussion
Vestibular/ocular-motor screening
Machine learning
Saccades
Vestibular-oculomotor reflex

ABSTRACT

The Vestibular/Ocular-Motor Screening (VOMS), an important component in acute (<72 h) sport-related concussion (SRC) assessment, is increasingly used alongside the Sport Concussion Assessment Tool (SCAT) and as part of the Military Acute Concussion Evaluation 2 (MACE2). VOMS demonstrates clinically useful diagnostic accuracy for acute SRC and improves the overall utility when added to the SCAT3. However, potential overlap among VOMS's vestibular and oculomotor items suggests the possibility of a more efficient version. VOMS and SCAT3 scores were analyzed for 3,958 preseason (47.8% female) and 496 acute-SRC (37.5% female) NCAA-DoD Concussion Assessment, Research, and Education (CARE) consortium collegiate athlete evaluations. Analyses revealed very large effect sizes (d = 2.39–2.45) and high correlations (rho = 0.95–0.99) among all VOMS items except near point of convergence distance (d = 0.79, rho ≤ 0.341). Receiver operating characteristic (ROC) curve analyses showed clinically useful discriminative utility for VOMS Total (AUC = 0.85) and the VOMS Total change score, where pretest symptoms were incorporated (AUC = 0.81). A modified VOMS (mVOMS) consisting of four items (smooth pursuits, horizontal saccades, horizontal vestibulo-ocular reflex, visual motion sensitivity) yielded identical AUCs to VOMS Total. Integer cutoff analyses suggest a score of ≥4 for VOMS Total and ≥4 for mVOMS Total optimizes concussion identification. Incorporating VOMS or mVOMS into SCAT3 (AUC = 0.79) significantly improved the combined tool's acute utility for acute concussion identification by a maximum of 4% (SCAT3+VOMS AUC = 0.84, SCAT3+mVOMS AUC = 0.83). Future versions of SCAT or MACE may want to consider incorporating a more parsimonious VOMS for the purpose of identifying acute concussion.

1. Introduction

Mild Traumatic brain injuries (mTBI) are a growing public health concern in both civilian and military settings. It is estimated sport-related concussions (SRCs) alone affect upwards of 3.8 million individuals annually in the United States (Baldwin et al., 2018; Cancelliere et al., 2017; Langlois et al., 2006), and over 354,913 United States military service members have sustained mTBIs in the line of duty since 2000 (DoD TBI Worldwide Numbers, n.d.). The true incidence of mTBIs is estimated to be nearly nine times higher, however, due to challenges surrounding detection and diagnosis (Baldwin et al., 2018; National Concussion Surveillance System, 2019). Clinicians often utilize multi-domain assessment tools when diagnosing acute concussion. One such tool, the Vestibular/Ocular-Motor Screening (VOMS), is a validated test designed to screen for vestibular and oculomotor symptoms and impairment following concussion (Elbin et al., 2018; Kontos et al., 2016). VOMS has also been shown to be an important component of acute (≤72 h) SRC assessments and is increasingly utilized concurrently with multi-domain concussion assessment tools including the Sport Concussion Assessment Tool (SCAT) and the Military Acute Concussion Evaluation 2 (MACE2) (Broglio et al., 2017; Kontos, Monti, et al., 2021; McCrory et al., 2017). Research has demonstrated VOMS to have clinically useful diagnostic accuracy for acute SRC in adolescent and collegiate-athlete populations (Ferris et al., 2021; Yorke et al., 2017) and

* Corresponding author at: Indiana University School of Optometry, 800 Atwater Ave, Bloomington, IN 47405, United States.
E-mail addresses: nport@indiana.edu, nlport@indiana.edu (N.L. Port).

https://doi.org/10.1016/j.visres.2022.108081
Received 16 August 2021; Received in revised form 7 March 2022; Accepted 7 June 2022
Available online 1 August 2022

to improve overall diagnostic power when added to the SCAT3 (Ferris et al., 2021). However, key questions remain, including the possible use of creating an abbreviated tool, the benefits of incorporating a modified version of VOMS within SCAT, the utility of the VOMS pretest total score, and the comparison of proposed clinical cutoff scores.

1.1. The potential for a modified, abbreviated VOMS

The first area of ambiguity surrounding the VOMS tool involves the potential overlap of vestibular and oculomotor tasks. VOMS includes eight oculomotor and vestibular items: Smooth Pursuits (horizontal and vertical), Horizontal Saccades, Vertical Saccades, Convergence, Near Point of Convergence distance (NPC, cm), Horizontal Vestibulo-Ocular Reflex (VOR), Vertical VOR, and Visual Motion Sensitivity (e.g., VOR cancellation). These tasks have repeatedly been shown to have a high internal consistency, with Cronbach's alphas ranging from 0.92 to 0.97 in both civilian (Moran et al., 2018) and military (Kontos, Monti, et al., 2021) populations. Time is often restricted in acute SRC evaluations, and especially so during competition when the rate of concussive injuries is highest (Baldwin et al., 2018). Abbreviating VOMS without sacrificing clinical utility would be beneficial to both patients and clinicians.

VOMS vestibular and oculomotor items have each demonstrated utility in acute concussion identification, supporting the idea that VOMS can be abbreviated without sacrificing clinical utility. Vestibular or oculomotor symptoms are common and reported in upwards of 65% of acute sport-related concussions (Ferris et al., 2021). Visual symptoms such as diplopia, blur, eyestrain, headaches, and difficulty tracking objects or reading are common in acute sport- (Gerberich et al., 1983) and non-sport related concussion (Greenwald et al., 2012), ranging from 51% to 83% of patients (Capó-Aponte et al., 2017; Ciuffreda et al., 2007; Ferris et al., 2021). Depending on the methods used to assess deficits associated with these symptoms, studies have found high prevalence of oculomotor-based disorders of eye movement (e.g., pursuit, saccades, fixation) ranging from 51% to 85% (Capó-Aponte et al., 2017; Ciuffreda et al., 2007; Lepore, 1995), convergence insufficiency from 12% to 36% (Capó-Aponte et al., 2017; Merezhinskaya et al., 2019), and accommodative dysfunction from 20% to 43% (Capó-Aponte et al., 2017; Ciuffreda et al., 2007; Merezhinskaya et al., 2019). Examinations of individual oculomotor skills have revealed further functional changes between concussed and non-concussed populations. Although small, reductions in gaze stabilization (Gottshall et al., 2003; McDevitt et al., 2016), increased lags and velocities for smooth pursuits (Heitger et al., 2004; Murray et al., 2020), and reduced saccadic accuracy, amplitude, velocity, and reaction-time tasks have been noted between concussed and matched non-concussed controls (Heitger et al., 2004; K. M. Kelly et al., 2019; Murray et al., 2020). Similar objective and subjective differences have also been noted for binocular oculomotor functions such as convergence and accommodation, although reductions in NPC and accommodative amplitude have displayed lower effect sizes between populations due to high variance amongst the two functions (Capó-Aponte et al., 2017; Ciuffreda et al., 2007; Green et al., 2010; McDevitt et al., 2016; Pearce et al., 2015).

The vestibular-driven VOMS tasks, which also require oculomotor function, follow similar trends, revealing increased symptoms after VOR testing and optokinetic stimulation (Elbin et al., 2018; Kontos, Eagle, et al., 2021; McDevitt et al., 2016). Indeed, when assessing symptom burden, individual VOMS items have shown acceptable diagnostic utility as shown by Area Under the Curve (AUC), ranging from 0.55 to 0.73 in youth populations (Elbin et al., 2021), and 0.51 to 0.90 in collegiate athletes (Kontos, Eagle, et al., 2021). Given the synergy between the oculomotor and vestibular systems, the high Cronbach's alphas, and the clinically useful AUCs for VOMS items, inherent redundancies may exist within the VOMS tool. In the context of acute concussion diagnosis where time is restricted (Al Jufali et al., 2015), a streamlined VOMS tool containing the minimal yet most effective items would give clinicians more time for evaluations while simultaneously providing high discriminability to aid the diagnosis of concussion.

1.2. Incorporating a modified version of VOMS into the SCAT

mTBIs often involve rapid-onset and multi-domain dysfunction, leading to continuously evolving somatic, cognitive, visual, vestibular, and emotional signs and symptoms. Furthermore, a lack of laboratory biomarkers and insufficient sensitivity of existing diagnostic imaging techniques force clinicians to rely on standardized, multi-domain symptom-based assessments (Buckley et al., 2015; Harmon et al., 2019; K. C. Kelly et al., 2014; McCrory et al., 2017). Two such multi-domain assessment tools are the Sport Concussion Assessment Tool (SCAT) (Echemendia et al., 2017), and the Military Acute Concussion Evaluation (MACE) (National Academies of Sciences, Engineering, and Medicine et al., 2019). Both tools include components evaluating somatic symptoms, cognitive function, and balance. The use of these multi-domain assessments improves discriminative utility for acute concussion compared to use of single-domain assessments alone (Ferris et al., 2021). The discriminability of VOMS in detecting vestibular and oculomotor dysfunction associated with acute mTBI led to the incorporation of VOMS into the second edition of the MACE (MACE2). A validation study comparing the diagnostic utility of VOMS to the SCAT3 battery (SCAT3 Symptom Evaluation, the Standardized Assessment of Concussion [SAC], and the Balance Error Scoring System [BESS]) demonstrated that incorporating VOMS into the existing SCAT3 increased the combined tool's sensitivity 7% to an AUC of 0.85 (Ferris et al., 2021) while adding <7 min to overall testing time. However, this analysis was conducted using the total VOMS symptom severity score (VOMS Total). Further investigation into the incorporation of any abbreviated VOMS tools into multi-domain assessments such as the SCAT could help identify possible improvements in diagnostic utility for clinicians with a minimal increase in assessment length.

1.3. Use of the VOMS pretest total score

The VOMS tool was developed to assess concussion symptoms through exacerbating tasks. Prior to any task, VOMS first assesses a summed pretest (e.g., pre-exacerbation) symptom severity score using an 11-point Likert scale across four domains: headache, fogginess, nausea, and dizziness. These domains are identical to those assessed within the SCAT3 Symptom Evaluation (Echemendia et al., 2017; Guskiewicz et al., 2013), except for the use of a 7-point scale, and are also included in the MACE2. Participants then complete oculomotor and vestibular tasks, and after each task, the four symptoms are reassessed. A total exacerbation symptom score (VOMS Total) can then be obtained by summing the individual post-task symptom scores. VOMS Total has shown clinically useful diagnostic utility in collegiate athletes, ranging from an AUC = 0.85 (Ferris et al., 2021) when using the raw total post-exacerbation symptom severity score alone (Kontos, Eagle, et al., 2021). However, a change score for VOMS Total, where the difference between total pretest symptoms and total post-exacerbation symptom has also been proposed and has demonstrated similar levels of predictive utility (AUC = 0.73) (Elbin et al., 2021). The presence of these two systems has led to confusion regarding how best to evaluate VOMS performance. As such, a direct comparison of the raw sum (VOMS Total) and the VOMS Total change score is needed.

1.4. Comparing VOMS clinical cutoff scores

A final issue involves calculating integer clinical cutoff scores for distinguishing concussed athletes from healthy controls. Clinical cutoffs are increasingly utilized in concussion assessments to quickly identify deviations from normative performance (Broglio et al., 2019; Caccese et al., 2020) and have been proposed for interpreting performance on the VOMS tool. Whereas each proposed cutoff method has yielded accuracy for acute concussion identification, there is a need for the

methods to be directly compared so clinicians can interpret which method(s) are most effective for their respective patient populations.

There remains ambiguity whether VOMS performance should be assessed by item, or by a single summed symptom severity score (VOMS Total). Two different itemized cutoffs have been proposed for use in adolescent populations, a third set has been proposed for use in collegiate athletes, and another set has been instituted for United States Service Members. Mucha and colleagues (2014) proposed cutoffs of mean NPC distance of ≥5 cm or a symptom provocation ≥2 on any VOMS items in youth (Mucha et al., 2014a). These cutoffs were also used by Yorke and colleagues (2017) (Yorke et al., 2017). In 2021, Elbin and colleagues' analyses of adolescent populations suggested a cutoff of ≥3 cm for mean NPC distance, and symptom cutoffs of ≥1 for any VOMS item (Elbin et al., 2021). In collegiate athletes, Kontos and colleagues (2021) suggest cutoffs of ≥4.5 cm for mean NPC distances or ≥1 for any VOMS item, except horizontal VOR (≥2) (Kontos, Eagle, et al., 2021). Finally, as of 2018, a fourth version of VOMS has been utilized to assess acute concussion status in US military personnel, with cutoffs of ≥5 cm for mean NPC distances or ≥1 for any VOMS item (National Academies of Sciences, Engineering, and Medicine et al., 2019). In addition to the itemized cutoffs proposed for VOMS, summed post-exacerbation symptom severity cutoffs (e.g., VOMS Total, or VOMS Overall) have also been suggested in both collegiate (≥8) and adolescent (≥3) athlete populations (Kontos et al. 2021; Elbin et al. 2021).The proposal of using total VOMS cutoffs, whether through raw sum or pretest change scores, is appealing as the method provides a path for avoiding potential multiple-comparison and Boolean analysis concerns. In the Mucha and colleagues (2014), Elbin and colleagues (2021), and Kontos and colleagues (2021) proposals, performance is considered suspect if an athlete exceeds the cutoff of any individual VOMS task (e.g., mean NPC distance OR Smooth Pursuit OR Horizontal Saccades, etc.). However, in all three manuscripts, the best discriminative utility of the tool is reported on a subset of VOMS items. These analyses raise the question as to whether the complete VOMS tool would have similar levels of predictive utility if all the individual items were considered using Boolean methodology, where deterministic relations between variables are evaluated (Romme, 1995). To date, the discriminability of VOMS tasks in US military personnel has not been evaluated, although test-rest findings have been recently published (Kontos et al. 2021).

Naturally, these differing cutoff scoring methods generate ambiguity over the choice of cutoff models to follow when assessing the VOMS. Given this, there is a need to directly compare proposed clinical cutoff methods to assess the impact various integer cutoffs, and cutoff scoring systems have on the discriminative utility of the VOMS tool in a large sample. Doing so could help clinicians determine which cutoff model if any, to employ during acute concussion assessments.

1.5. Optimizing VOMS

To address the four issues raised above, rigorous statistical techniques were employed to analyze data from a large sample of collegiate student-athletes from the National Athletic Association-Department of Defense (NCAA-DoD) Concussion Assessment, Research and Education (CARE) Consortium. First, given the synergy between the oculomotor and vestibular systems and the high AUCs and Cronbach's alpha for each item, we hypothesized that a shortened VOMS would have equal discriminative ability to detect acute SRC (Aim 1). Secondly, we hypothesized that incorporating either the current or a modified (shorter) version of the VOMS into the SCAT would improve SCAT performance (Aim 2). Third, based on the predictive utility of SCAT3 graded symptom scores and VOMS Total post-exacerbation scores alone, we hypothesized that incorporating the pretest symptoms into VOMS Total does not improve diagnostic accuracy (Aim 3). Finally, we hypothesized that using a dataset of this size it would be possible to compare clinically meaningful integer cutoff scores for the VOMS (Aim 4).

2. Materials and methods

Complete methods on participant recruitment, enrollment, and assessment for the multi-site CARE Consortium have previously been published (Broglio et al., 2017). At all sites, NCAA varsity-level student-athletes and military service academy cadets were enrolled after providing informed consent. IRB approval was obtained by individual institutional review boards and the US Army Human Research Protection Office. All study activities were congruent with the Declaration of Helsinki.

2.1. Participants

Of the 34,709 unique participants in the CARE dataset as of February 23, 2021, 3,464 individuals from seven NCAA Division-I, two Division-II, and two Division-III institutions completed VOMS and SCAT3 assessments and were screened for inclusion into the current investigation: Indiana University, University of Pittsburgh, University of California Los Angeles, University of Delaware, University of North Carolina, Wake Forest University, University of Wisconsin-Madison, West Chester University, Winston-Salem State University, and Humboldt State University. Assessments were conducted at athletes' preseason evaluations or at any acute (≤6-48 hrs) concussion timepoint. Athletes with a history of vestibular disorders, moderate/severe brain injury, and/or brain surgery were excluded (n = 20), yielding a final cohort of 3,444 athletes and 3,958 preseason evaluation encounters. Four hundred and fifty-four athletes sustained at least one acute concussion during the study, 38 experienced two concussions, and 2 experienced three concussions. The final dataset therefore included 496 acute concussion evaluations (Fig. 1). Data from this sample has previously been analyzed by our research group (Ferris et al., 2021, 2022).

2.2. Assessment variables

2.2.1. Vestibular/ocular-motor screening

The VOMS tool is a free, quick, pen-and-paper assessment of vestibular and oculomotor symptoms through nine items (Kontos, Eagle, et al., 2021). Designed for use with minimal training, items include (1) Pretest Symptoms, (2) Smooth Pursuit, (3) Horizontal Saccades, (4) Vertical Saccades, (5) Convergence, (6) Near Point of Convergence (NPC), (7) Horizontal VOR, (8) Vertical VOR, and (9) Visual Motion Sensitivity (VMS). At the conclusion, the sum of all the symptom severity scores is recorded for the VOMS Total score. Patients are asked to report pretest symptoms on an 11-point Likert scale (0–10) across four categories: headache, nausea, fogginess, and dizziness prior to any VOMS task. Symptoms are then reassessed at the conclusion of each task to determine whether the vestibular/oculomotor activity exacerbated symptoms. Symptom severity can be summed individually per task (max = 40) (Mucha et al., 2014a), or summed across all exacerbating tasks, providing an overall total post-exacerbation symptom score (VOMS Total, max = 280) (Ferris et al., 2021; Kontos, Eagle, et al., 2021). Scores can also be evaluated in terms of change scores from pretest for individual items (Elbin et al., 2021; Yorke et al., 2017) or the overall VOMS symptom severity score (Elbin et al., 2021). Of note, NPC distance is measured from the tip of the nose with a 14-pt target, and not from the plane of the eyes as is standard in ophthalmic evaluations (Cochrane, 2017).

The architecture of the tool results in two different scoring arms for VOMS: 1) symptom severity (either individual items or total score), and 2) mean NPC distance. It also results in two different scoring methodologies for the symptom severities: 1) raw sum, and 2) change scores. All published VOMS scoring methods were evaluated in this study. As the term baseline can be easily confused with the preseason baseline testing commonly performed by many sports medicine programs, the VOMS "Baseline" symptoms term, proposed by Mucha and colleagues (2014), is referred to here as pretest, as recently suggested (Elbin et al., 2018,

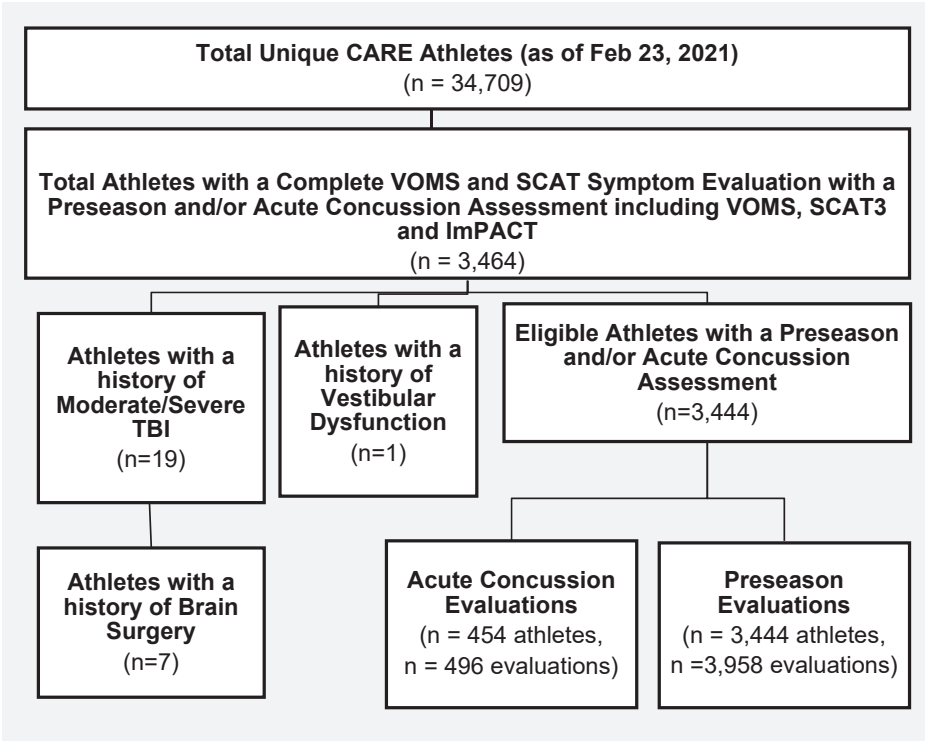

Fig. 1. Athlete Selection Schematics.

2021). VOMS symptom severity scores are reported as raw preseason or post-injury scores, unless otherwise specified as change scores from pretest to post-exacerbation.

2.2.2. Sport concussion assessment tool, 3rd edition

The SCAT3 is a standard multidimensional assessment of the physical, neurological, symptomatic, cognitive, and spino-vestibular systems (Guskiewicz et al., 2013). Currently in its 5th edition, SCAT3 was the most current edition at the beginning of CARE and was selected for use throughout the duration of the study. It contains multiple independent subscales, including the modified Balance Error Scoring System (mBESS), the standardized assessment of concussion (SAC), and a graded symptom evaluation, summarized below.

The mBESS assesses the spino-vestibular system. Participants conduct 3 single and tandem-leg stances barefoot (or with socks) on a firm surface with eyes closed for 20 s each (Guskiewicz et al., 2013). Scoring begins at zero and increases for each balance deviation, up to a total of 10 deviations per stance (max = 30). CARE protocols included both modified and full BESS (mBESS followed by a repeat of all three stances on an AIREX™ foam pad) (Broglio et al., 2017), but the standard SCAT3 includes only mBESS. Consequently, only mBESS scores were evaluated in this study.

The SAC evaluates neurocognition including orientation, immediate and delayed recall, and concentration evaluations (Guskiewicz et al., 2013). Scores can be calculated either by component or as the total sum of all components. While a 10-word list is now incorporated into the SCAT5, (Echemendia et al., 2017) the 5-word list was standard at the time of CARE. The SAC is reverse-scored, with a perfect performance of 30 points and points taken away for errors. However, this study's ROC analyses (see below) were calculated using an inverted and absolute SAC score (e.g., where scores began at 0 and increased with each error to a maximum of 30).

The final subscale of SCAT3 assessed in this study was the SCAT3 Symptom Evaluation, a 22-item neurological and somatic questionnaire on a 7-point Likert scale (0–6). The SCAT3 Symptom Evaluation was included in the recently updated SCAT5 without any changes (Echemendia et al., 2017). Items include common symptoms such as headache, nausea, fogginess and dizziness, as well as balance problems, sensitivity to light, and blurred vision. The symptom evaluation can be scored as either the total number of items with a positive symptom report (max = 22), or as the sum of the severity of all symptoms (max = 132). For this investigation, the total symptom severity score was used.

2.3. Statistical analysis

Analyses examined relationships between preseason and acute post-injury performance with regard to VOMS items, and between VOMS and the SCAT3 battery (SCAT3 Symptom Evaluation, SAC, mBESS) and were conducted using custom scripts in R (R Core Team, Vienna, Austria) (R Core Team, 2017; Wickham, 2016) and MATLAB (MathWorks, Natick, MA) across three distinct phases for each aim of the study. First, descriptive statistics assessing raw preseason and acute post-injury scores were calculated. Second, relationships between variables were determined through multiple methods (see below). Third, predictive power for variables was calculated and compared using ROC analyses.

2.3.1. General statistical methods

Standard descriptives included the mean and variance between timepoints for each tool. As previous research has noted significant skew for VOMS items and SCAT3 subscales, medians and 95th percent confidence intervals (CI) were also calculated. Reported median and 95% CI results were rounded to the nearest integer for all ordinal Likert scale assessments. The Kolmogorov-Smirnov test was utilized to calculate statistical significance between concussed and non-concussed athletes, and effect sizes were calculated using Cohen's d, where $d \leq 0.40$ was considered minimal, $0.40 < d \leq 0.80$ moderate, $d > 0.80$ large, and $d > 2.00$ very large (Bakker et al., 2019; Ialongo, 2016; Sawilowsky, 2009). Spearman's rho was utilized to determine the preseason and acute SRC correlations between VOMS items, and Deming regressions (Deming, 1943) were utilized to determine the linear relationships between pairs of dependent VOMS and SCAT variables (Linnet, 1993; Marín-Franch et al., 2013). Additional specific methods used for each of the three aims

are detailed below.

Once general descriptives and regression-based analyses were complete, ROC curve analyses were conducted to determine predictive utility. Youden's method was utilized to select the best indicator of discriminative utility (e.g., AUC), and Delong's method was used to compare AUCs and determine significance (DeLong et al., 1988; Perkins, 2004). AUCs range from zero to one, with values above 0.50 providing the diagnostic probability of a test correctly identifying an individual with a specific condition from a healthy individual (Massof & Emmel, 1987). As such, any AUC $\leq$ 0.60 was considered very poor for positively identifying concussion. AUC values from 0.60 to 0.80 were considered moderate, and AUCs $>$ 0.80 were considered high and clinically useful (Mandrekar, 2010). All analyses utilized an *a priori* type I error rate of $p < 0.01$.

2.3.2. Aim 1: Determine whether the current VOMS tool can be streamlined without reducing diagnostic accuracy

To address this question, each VOMS item was individually analyzed. Descriptive statistics for each item were calculated, and correlations between items were analyzed at the preseason and acute-SRC time frames. First, logistic regression was conducted to evaluate whether any particular item or grouping of items significantly contributed to the performance of the overall tool. Following this, forward, backward, and bidirectional stepwise logistic regressions were conducted to identify the highest contributing subset of VOMS items. Multiple criterion methods were compared when using these models, including maximizing the Akaike information criterion (AIC), the Bayesian information criterion (BIC), the R-square statistic, and the adjusted R-square statistic (StepReg: Stepwise Regression Analysis, n.d.; Venables & Ripley, 2002).

The next technique used was principal component analysis (PCA) (Kassambara, 2017; Lê et al., 2008). This type of analysis is ideal for highly correlated data as it seeks to summarize the impact individual variables have on the principal components of the model. Finally, a brute-force, all-possible combinations method was conducted using a custom support-vector machine learning model in MATLAB derived from the Machine Learning Toolbox (Paluszek & Thomas, 2020).This method used ROC analyses and compared the AUC of each individual VOMS item. Once individual AUCs were calculated, the method was repeated for every possible combination of VOMS items to determine the resultant AUCs for all-possible single, double, triple, quadruple, quintuple, sextuple, septuple, and octuple combinations of VOMS items.

2.3.3. Aim 2: Determine whether incorporating a streamlined version of VOMS into SCAT3 improves diagnostic accuracy

Upon completion of potential data-reduction analyses, a modified VOMS (mVOMS) model was selected. Descriptive statistics and Deming regressions were then conducted for mVOMS in relation to both VOMS Total and the SCAT3 as noted above. The linear relationship between the SCAT3 Symptom Evaluation and mVOMS Total revealed by Deming regression was then used to create a scaled version of mVOMS, set on a 0–6 Likert scale to match that of the SCAT3 Symptom Evaluation. Finally, ROC analyses were performed to determine: 1) the relative predictive utility between VOMS Total, mVOMS, mVOMS scaled, and SCAT3; and 2) how these utilities changed if any variation of VOMS was incorporated into the existing SCAT3 battery.

2.3.4. Aim 3: Determine whether incorporating pretest symptoms into the VOMS Total improves diagnostic accuracy

To determine whether incorporating pretest symptoms into VOMS via a pretest change score improves diagnostic accuracy, the post-exacerbation symptom severity scores for each of the seven symptom-based items were subtracted from the VOMS Pretest total symptom severity score and then summed to create a VOMS Total Change Score. This change score was used to generate and analyze descriptive statistics between VOMS Total, VOMS Pretest Total, and the VOMS Total Change Score. Deming regression was then utilized to examine the relationship between VOMS Pretest Total (max = 40) and VOMS Total (max = 280). To permit this direct evaluation, the average VOMS Total symptom severity score was utilized, by dividing VOMS Total by seven to account for the seven VOMS items with symptom severity reports. Finally, ROC analyses were conducted, comparing the predictive utility of the existing VOMS Total, VOMS Pretest Total, and VOMS Total Change Score. Results are reported as best AUC, along with best sensitivity and specificity as determined by Youden's method (Bewick et al., 2004).

2.3.5. Aim 4: Using AUC, sensitivity, and specificity, compare clinically meaningful cutoffs for VOMS and mVOMS

The AUCs, sensitivities, and specificities from Aims 1 through 3 were calculated using raw preseason and post-injury scores for VOMS (e.g., VOMS Total symptom severity scores ranging from 0 to 280). This methodology results in a relatively smooth ROC curve because all possible cutoffs for a given variable is utilized, which produces a range of sensitivity and specificity values. Conversely, when single integer cutoffs are used ROC curves transform into two line-segments, resulting in reductions in the area under the curve. One of the goals of Aim 4 is to directly compare already published clinical cutoffs with the cutoffs calculated in this study. ROC analyses were conducted between the 496 concussed evaluations and the 3,958 baseline evaluations to determine and compare the impact different VOMS items, VOMS Total, and mVOMS Total integer cutoffs have on the AUC, sensitivity, and specificity. A large range of integer cutoffs (1–10) were evaluated for both methods. As the evaluated clinical integer cutoffs fall along the all-possible raw score AUC curve for each method, cutoffs are displayed as points along the raw score AUC curve for VOMS Total (Fig. 11A) and mVOMS Total (Fig. 11B). The discriminability for these cutoffs is reported in terms of sensitivity and specificity, and Youden's index. An optimal clinical cutoff was identified when there was not a significant improvement in the AUC, sensitivity and specificity as determined by Delong's method ($p < 0.01$).

The same methodology was utilized to compare boolean cutoffs for individual VOMS items (Elbin et al., 2021; Mucha et al., 2014a; Yorke et al., 2017) and integer cutoffs for VOMS Total (Elbin et al., 2021; Kontos, Eagle, et al., 2021), including the optimal integer cutoffs for VOMS Total and mVOMS Total. As these methods utilized different scoring systems, their discriminability had the potential to fall along different all-possible AUC curves. Given this, discriminability performance for each method is displayed as their respective two-line segment ROC curves, and AUC, sensitivity, specificity, and Youden's J index (sensitivity + specificity-1) is reported for each method (Bewick et al., 2004). Significance was reported as determined by Delong's method ($p < 0.01$) and reflected a change in discriminability for each respective method as compared to the optimal VOMS Total cutoff.

3. Results

3.1. Population demographics

A total of 3,958 athlete evaluations (1,889 female; 2,069 male) were included in the study (Table 1). This data set has been evaluated by this group (Ferris et al., 2021, 2022), and as previously reported, there were no significant impacts of sex on demographics or raw discriminative utility of VOMS or SCAT components in this data set; therefore, the sexes were combined in this analysis. Visual and/or balance symptoms were reported by two-thirds (66.1%) of the acutely concussed athletes at their acute evaluation (as noted by a SCAT3 symptom severity score $>$ 0 for "sensitivity to light", "blurred vision", "dizziness", or "balance problems"). Three subjects (0.60%) had positive CT findings for facial fractures at the time of injury but had no other structural/neurological findings. Inclusion or exclusion of these subjects had no impact on the results of any analysis. As such, they were retained in this investigation.

Table 1
Population Demographics.

	Total Count (% total)
Preseason Demographics	
Total Preseason Evaluations	3958 (47.8% female)
Mean Age (SD)	19.44 (1.48)
Prior Concussions at Baseline	
0	2479 (62.63%)
1	673 (17.00%)
2	177 (4.47%)
3+	67 (1.69%)
N/A	48 (1.21%)
Acute SRC Demographics	
Acute Concussion Evaluations	496 (37.5% female)
Loss of Consciousness	20 (4.03%)
Retrograde or Post-Traumatic Amnesia	47 (9.47%)
Visual Symptoms	285 (57.50%)
Sensitivity to light	*263 (53.02%)*
Blurred vision	*141 (28.43%)*
Balance Symptoms	246 (49.60%)
Dizziness	*227 (45.77%)*
Balance problems	*158 (31.85%)*

3.2. Aim 1: Determine whether the current VOMS tool can be streamlined without reducing diagnostic accuracy

Stair histograms and descriptive statistical summaries for VOMS items are shown in Fig. 2, panels A-I. All items demonstrated significant increases after a concussion ($p < 0.01$) and had very large effect sizes ($d \geq 2.34$) with the exception of mean NPC ($d = 0.79$). Individual VOMS post-exacerbation items had a median value of 0 and increased to a median of 5 post-concussion, whereas VOMS Pretest rose from 0 to 4. Mean NPC distance increased from a median of 2.33 cm to 4 cm but demonstrated a wider 95th percentile confidence interval (3.70–4.00) than the other VOMS items.

Preseason and acute-SRC correlations between VOMS items revealed high correlations between all symptom-based items at both non-concussed and concussed timeframes ($p < 0.01$) (Fig. 3). Preseason and acute-SRC correlations were much lower for any symptomatic item and mean NPC distance, with preseason correlations between mean NPC and Pretest, Smooth Pursuit, and Vertical Saccades not reaching significance ($p > 0.01$).

Given the similar effect sizes and high correlations between VOMS items, multiple statistical methods were explored to analytically reduce the length of the VOMS tool. Logistic regression was first examined as it has been utilized to evaluate the relative contribution of VOMS items to the tool. All items had similar and small coefficient estimates ($\beta \leq 0.23$) (Table 2).

Both forward and backward stepwise logistic regressions revealed inconsistencies with VOMS item retention based upon multiple selection criterion used (Akaike information criterion, Bayesian information criterion, R-squared, etc.), ranging from retaining one VOMS item, to 8. For any model, retained VOMS item's coefficients remained minimal ($\beta \leq 0.27$).

Next, principal component analysis attempted to discern whether any VOMS item, or group of items, contributed more to the principal components of VOMS along two dimensions. Both preseason (Fig. 4A) and acute-SRC (Fig. 4B) results were similar and echoed other data-reduction methods. All VOMS items contributed similarly to the overall principal components, as shown by a similar distance from the origin for each item's vector. The symptomatic items of VOMS grouped tightly together along dimension 1 and contributed the most to the tool at both the preseason (80.4%) and acute-SRC (88.3%) timeframes. Mean NPC displayed similar strength of contribution to the secondary, orthogonal dimension (preseason 11.1%, acute-SRC 9.9%). No single item or group of items stood apart as more contributive than others at either time frame.

After regression and PCA analyses were complete and revealed redundancies, ROC analyses on the 496 concussed and 3,958 non-concussed athlete evaluations were conducted using post-exacerbation VOMS performance scores on all possible combinations of VOMS items. Individual VOMS items displayed clinically useful diagnostic utilities for acute SRC detection with an average AUC = 0.85 (Fig. 5). Mean NPC distance was less discriminative, with an AUC = 0.64. These levels of predictive power were maintained for pairs of any variable and did not significantly improve beyond triplet- or quadruple-combinations of variables, peaking at AUC = 0.85. The mean NPC distance, horizontal VOR and VMS each demonstrated a significant difference in discriminative utility from VOMS Total ($p < 0.01$).

3.3. Aim 2: Determine whether incorporating a streamlined version of VOMS into SCAT3 improves diagnostic accuracy

Aim 1 demonstrated high redundancies between VOMS items, and statistical-based methods of variable selection failed to reveal a clear solution to reducing the VOMS tool. Therefore, a more streamlined, modified VOMS was created based upon the following considerations: the overall discriminability (Cohen's d) of each respective tool, the presence of repeat (e.g., horizontal, and vertical) tasks, and the relative ease of task conduction/patient education. Mean NPC distance demonstrated a poor ability to discriminate between concussed and non-concussed ($d = 0.79$) relative to the other items, and both saccades and VOR are duplicated along horizontal and vertical meridians, with similar effect sizes for both. Finally, convergence had similar effect sizes as the VMS and VOR items but did not add additional information beyond what would have been provided by either alone. Taking all these aspects into consideration, our modified VOMS (mVOMS) contains: 1) Smooth Pursuit, 2) Horizontal Saccades, 3) HVOR, and 4) VMS. Of note, and in line with the original design of the tool, this model retains both primary oculomotor and primary vestibular-driven tasks.

Descriptive statistics and histogram distributions of mVOMS Total compared to the current VOMS Total and the SCAT3 Symptom Evaluation severity score are shown in Fig. 6. All three assessments demonstrated significant differences between the preseason and acute-SRC timeframes ($p < 0.01$). mVOMS Total and VOMS Total show identical ($d = 2.44$) and very large effect sizes that are slightly higher than that for SCAT3 Symptom Severity ($d = 1.74$). Whereas the non-parametric exponential decay shape was retained for mVOMS, it showed better agreement between its mean (28.81) and median (21) than did VOMS Total.

High, significant, correlations between mVOMS Total and VOMS Total at both the preseason ($rho_{pre} = 0.98$) and acute-SRC ($rho_{acute} = 0.99$) timeframes ($p < 0.01$) were also seen (Fig. 7). Both mVOMS Total ($rho_{acute} = 0.87$) and VOMS Total ($rho_{acute} = 0.87$) displayed similar relationships with SCAT3 Symptom Severity and demonstrated a systematic increase in symptom severity at the acute-SRC time frame relative to SCAT. This relationship held when mVOMS was scaled to a 0–6 Likert scale to match that of the SCAT3 Symptom Severity using the calculated acute Deming regression formula (Fig. 7D). The scaled mVOMS Total demonstrated high correlation with SCAT ($rho_{acute} = 0.87$), falling close to the unity line and indicating an accurate transformation between the two Likert scales. This is further emphasized by the continued high discriminability of the scaled mVOMS Total, with the Cohen's *d* remaining very large at 2.44 (Fig. 7E).

When mVOMS Total (AUC = 0.85), mVOMS Total scaled to the SCAT Likert scale (AUC = 0.85) and VOMS Total (AUC = 0.85) were analyzed for predictive utility, all three showed clinically useful levels of diagnostic accuracy, with near-identical sensitivities (0.77) and specificities (0.83), and no significant differences noted between methods ($p = 0.2125$) (Fig. 8A). While still discriminative, the SCAT3 battery (SCAT3 Symptom Evaluation, SAC, mBESS) demonstrated lower utility (AUC =

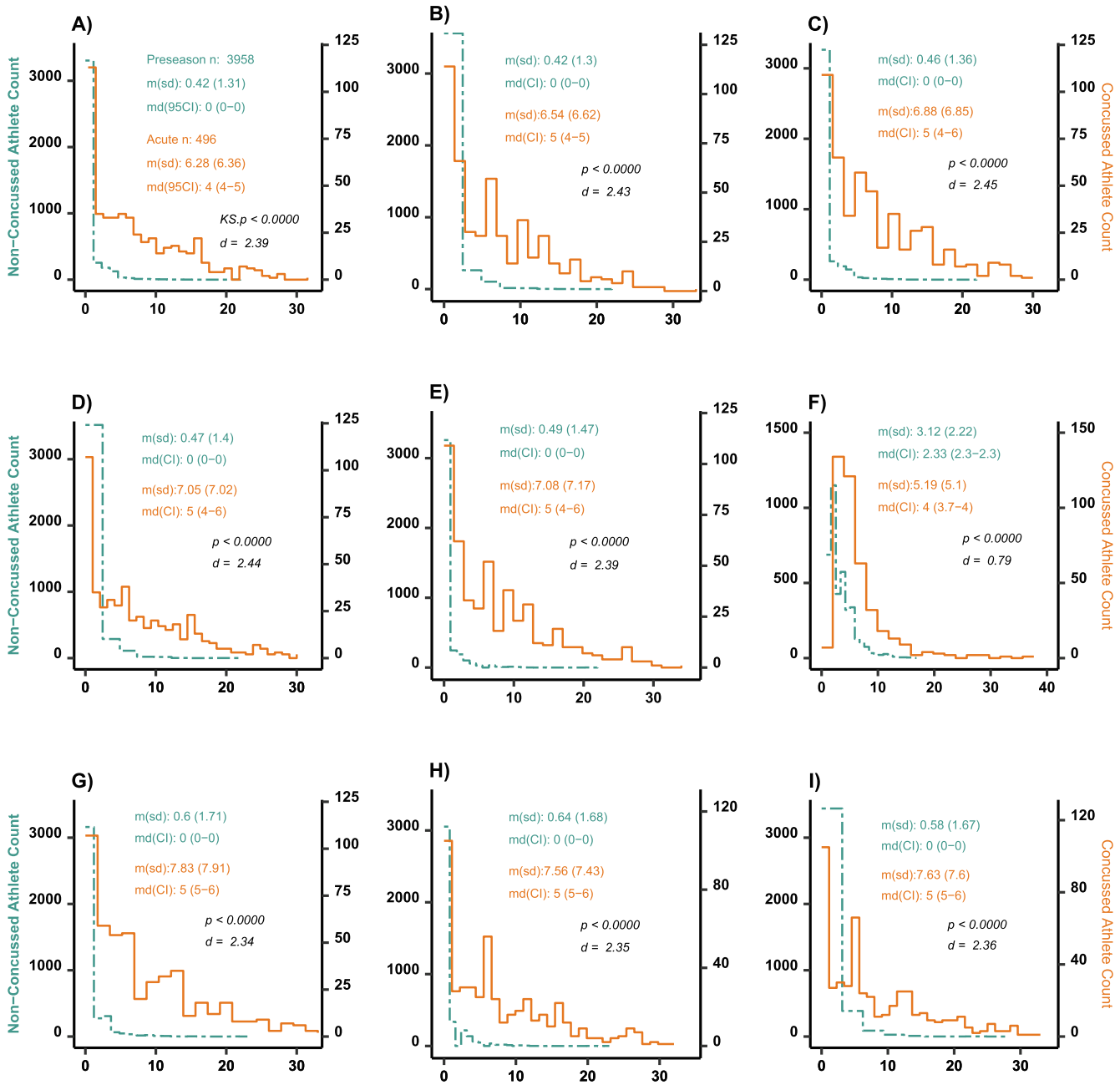

Fig. 2. Distributions and Descriptive Statistics for VOMS Items. Preseason non-concussed (teal) and acutely concussed (orange) scores for: A) VOMS Pretest, B) Smooth Pursuit, C) Horizontal Saccades (HSac), D) Vertical Saccades (VSac), E) Convergence, F) Mean NPC, G) Visual Motion Sensitivity (VMS), H) Horizontal VOR (HVOR), and I) Vertical VOR (VVOR) symptom severity scores. The left axis displays the count of non-concussed athletes, while the right axis relays the count of concussed athletes. The notation "m" refers to the mean whereas "md" refers to the median. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

0.79) relative to the VOMS tools. This lower utility is driven by the higher preseason mean and variance, and smaller effect size seen for SCAT3 in Fig. 6. When any of the three VOMS variations were incorporated into the SCAT3 battery, significant increases in diagnostic utility for the combined SCAT tool were demonstrated ($p < 0.01$), with SCAT3 + VOMS Total (AUC = 0.84), SCAT3 + mVOMS Total (AUC = 0.83), and SCAT3 + scaled mVOMS Total (AUC = 0.83) all increasing the overall accuracy by a similar amount (Fig. 8B). Incorporating the VOMS Pretest alone into SCAT3 also improved the predictive power (AUC = 0.81), although to a lesser extent than the VOMS post-exacerbation scores. Overall, the differences between each variation of the incorporated VOMS were not significant ($p > 0.01$).

3.4. Aim 3: Determine whether incorporating pretest symptoms into the VOMS Total improves diagnostic accuracy

Fig. 9, panels A-C, display histogram distributions and summarize the descriptive statistics for VOMS Total, VOMS Pretest and VOMS Total change score. Both VOMS Total ($d = 2.44$) and VOMS Pretest ($d = 2.39$) demonstrated very large effect sizes and significant increases following concussion ($p < 0.01$). The VOMS Total change score displayed significant differences ($p < 0.01$) between preseason and acutely concussed timepoints and provided similar discriminability between time frames ($d = 2.43$). Due to the non-parametric nature of the distributions, large differences at both the preseason and acute-SRC timeframe were noted between the means and medians for all three variables. For example,

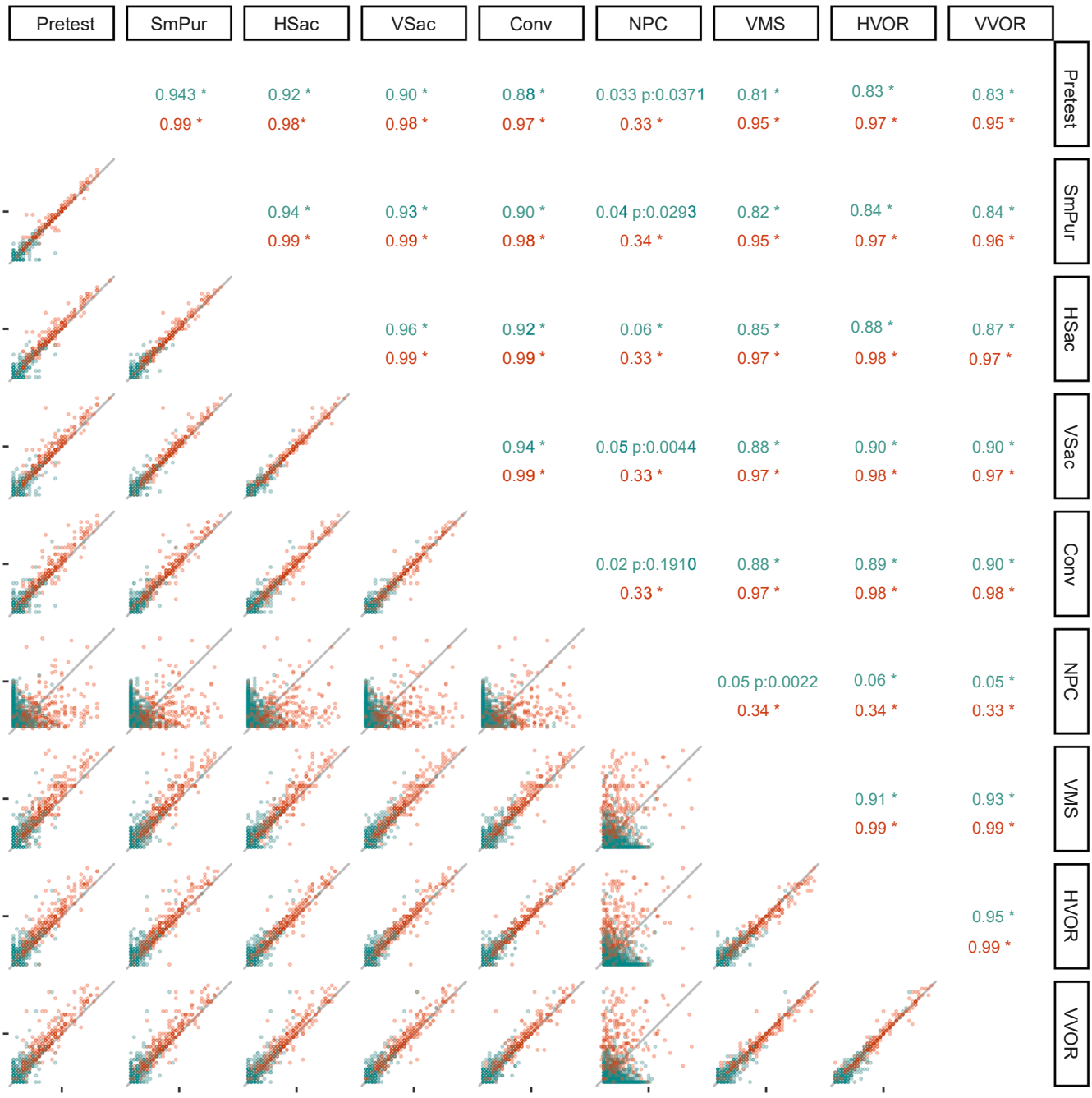

Fig. 3. Preseason and Acute-SRC Correlations between VOMS Items. Preseason non-concussed (teal) and acutely concussed (orange) scores (lower left triangle), and correlation coefficients and statistical significance (upper right triangle) for: VOMS Pretest Smooth Pursuits (SmPur), Horizontal Saccades (HSac), Vertical Saccades (VSac), Convergence (Conv), Mean NPC (NPC), Visual Motion Sensitivity (VMS), Horizontal VOR (HVOR), and Vertical VOR (VVOR) symptom severity scores. Unity lines are denoted as the light grey line. Both X and Y axes range from 0 to 30 with tick marks located at 15. Asterisks denote statistical significance at $p \leq 0.01$. Where not significant, the resultant p value of a correlation at either the non-concussed (teal) or acutely-concussed (orange) timeframe is given next to its respective correlation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

whereas the mean (m) VOMS Total score post-concussion was 50.57, the median (md) was 36.

Deming regressions (Fig. 10A) revealed a linear relationship between VOMS Pretest and Average VOMS Total. Correlations were high at both the preseason (rho = 0.77) and acute-SRC (rho = 0.98) timeframes ($p < 0.01$), with a high concentration of symptom scores at zero for both variables at the preseason baseline. Of note, with the symptom scales matched, VOMS Post-Exacerbation demonstrated a systematic uptick in symptoms over Pretest at both the preseason and acute-SRC time frames. For the acute SRC phase, this linear relationship revealed approximately an 11% increase in symptom severity scores post-exacerbation.

When the predictive utility of the three variables was compared (Fig. 10B), VOMS Pretest and VOMS Total Change Score showed high diagnostic accuracy on par with VOMS Total (AUC = 0.85 for all). There was no significant difference to the diagnostic utility of any variable.

3.5. *Aim 4: Using AUC, sensitivity, and specificity, compare clinically meaningful cutoffs for VOMS and mVOMS*

ROC analyses were conducted to evaluate potential cutoff scores for

Table 2
Logistic Regression for VOMS Items.

VOMS Item	β Estimate	Standard Error	t value	p value
Intercept	−3.61	0.12	−30.92	0.0000
Pretest	−0.038	0.11	−0.34	0.7317
Smooth Pursuits	0.23	0.13	1.75	0.0798
Horizontal Saccades	0.14	0.14	0.94	0.3455
Vertical Saccades	0.17	0.14	1.20	0.2295
Convergence	−0.01	0.99	−0.09	0.9298
Mean NPC Distance	0.09	0.02	4.13	0.0000
VMS	−0.01	0.08	−0.07	0.9418
Horizontal VOR	−0.14	0.11	−1.30	0.1949
Vertical VOR	0.16	0.11	1.50	0.0134

VOMS Total (Fig. 11A) relative to mVOMS Total (Fig. 11B) and to compare their discriminative utilities to previously proposed VOMS clinical cutoffs (Fig. 12). As noted in the methods, optimal clinical cutoffs were identified when there was not a significant improvement in the AUC as determined by Delong's method ($p < 0.01$). Note, the AUCs in Fig. 12 are different from above due to the binomial method of cutoffs. ROC curves are restricted to two-line segments which result in lower AUCs.

For VOMS Total and mVOMS Total potential cutoffs, each successive cutoff integer was compared to the previous one to determine if there was a significant increase in discriminative utility or not. It should be emphasized that integer cutoffs for VOMS Total and mVOMS Total are single points selected from the all-possible options seen when using raw scores. As such, these single integer cutoffs all fall upon the raw ROC curve for these respective measures. Whereas the sensitivity and specificity varied between potential cutoffs of 4, 5, and 6 for VOMS Total, there were no significant differences beyond a cutoff of ≥4 for VOMS Total, which demonstrated an AUC of 0.80. Similarly, mVOMS Total demonstrated no significant differences in AUC for cutoff scores ≥4 (AUC = 0.80). For comparison with the literature Youden's index is also reported. In general, increasing cutoffs beyond 4 for both VOMS Total and mVOMS Total increased specificity at the expense of sensitivity.

Multiple VOMS methods and cutoffs have been proposed. To date, the discriminative utility of these methods has not been directly compared within the same data set. As such, all VOMS methods cutoffs were compared to a VOMS Total cutoff of ≥4 as proposed above to determine if there was a significant increase in discriminative utility or not (Fig. 12). Evaluating VOMS items independently, whether by total raw sum), or by change score resulted in significant reductions in AUC, sensitivity, and specificity from any proposed VOMS total symptom severity cutoff method ($p < 0.01$). There was no significant difference in diagnostic utility between VOMS Total ≥ 4 (AUC = 0.80), mVOMS Total ≥ 4 (AUC = 0.80) or VOMS Overall ≥ 8 (AUC = 0.80) ($p > 0.01$). However, the sensitivities and specificities of these respective methods did vary, with VOMS Total ≥ 4 (78%), mVOMS Total ≥ 4 (77%) and the overall VOMS change score ≥3 (79%) demonstrating higher sensitivities than VOMS Overall ≥ 8 (73%). VOMS Overall ≥ 8 demonstrated the highest specificity, at 86%.

4. Discussion

The VOMS tool is one of the few free, quick, pen-and-paper vestibular/oculomotor assessments available for concussion screening. VOMS has high accuracy for acute concussion identification. However, individual VOMS items demonstrate high correlations, similar effect sizes, and AUCs to each other, suggesting these oculomotor and vestibular tasks contain significant redundancies. As shown through ROC analyses, streamlining VOMS to a maximum of four vestibular and oculomotor tasks (Smooth Pursuits, Horizontal Saccades, Horizontal VOR, and VMS) and assessing the total sum of those post-exacerbation symptom severities provides the same diagnostic utility as the full VOMS tool. Moreover, the addition of this shorter, modified VOMS to the SCAT3 boosts the overall discriminative utility of the existing SCAT3 battery (SCAT3 Symptom Evaluation, SAC, mBESS). Finally, ROC analyses highlight how Boolean integer cutoffs of individual VOMS items impact the tool's discriminative utility, and how utilizing cutoffs to assess total symptom score provides greater diagnostic utility than assessing each item individually. Together, these findings reinforce the clinical utility of the VOMS tool for acute concussion assessment and suggest pathways for optimizing the VOMS tool beyond its current published format.

VOMS works well as a screening tool for vestibular and oculomotor dysfunction. These two systems have high functional overlap such that normal performance of either requires synergistic interaction (Leigh & Zee, 2015). However, as shown in the results, this synergy also leads to potential redundancies. All individual VOMS items except mean NPC distance correlate highly and possess discriminability for acute concussions. This congruence complicated data reduction methods because items consistently demonstrated nearly equal contributions to the overall tool. This is shown best by the individual and all-possible combination ROC analysis (Fig. 5), where each individual symptom-based VOMS item demonstrated similar predictive power (average AUC = 0.85) and any combination of four or more VOMS items yielded similar overall power compared to VOMS Total.

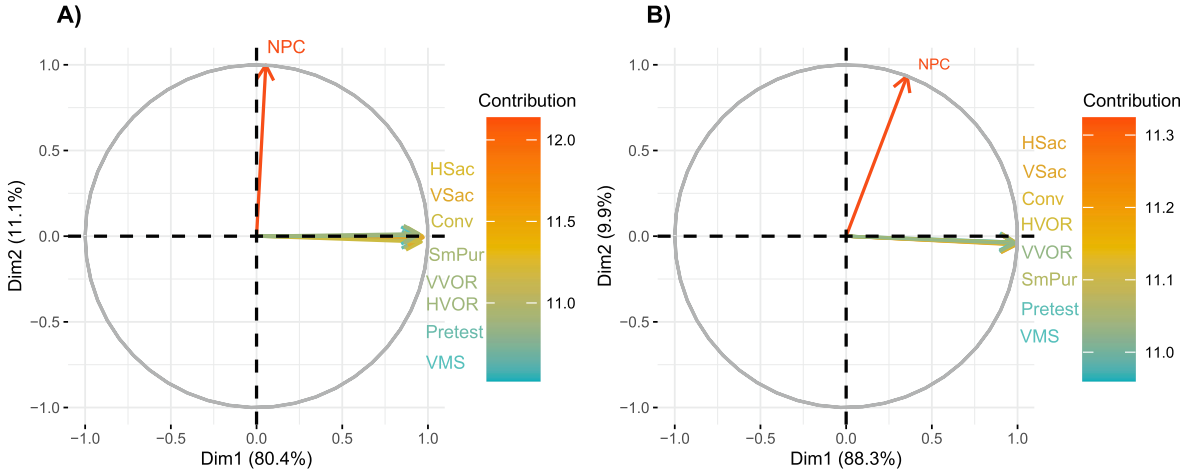

Fig. 4. Principal Component Analysis for VOMS Items. A) Preseason (non-concussed) and B) acutely concussed principal component analysis results for VOMS Pretest (Pretest), Smooth Pursuit (SmPur), Horizontal Saccades (HSac), Vertical Saccades (VSac), Convergence (Conv), Mean NPC (NPC), Horizontal VOR (HVOR), Vertical VOR (VVOR) and Visual Motion Sensitivity (VMS). Radial plots show the relative contribution of each item along the two principal components (Dim1, Dim2), and a heatbar categorizes the relative strength of each individual item.

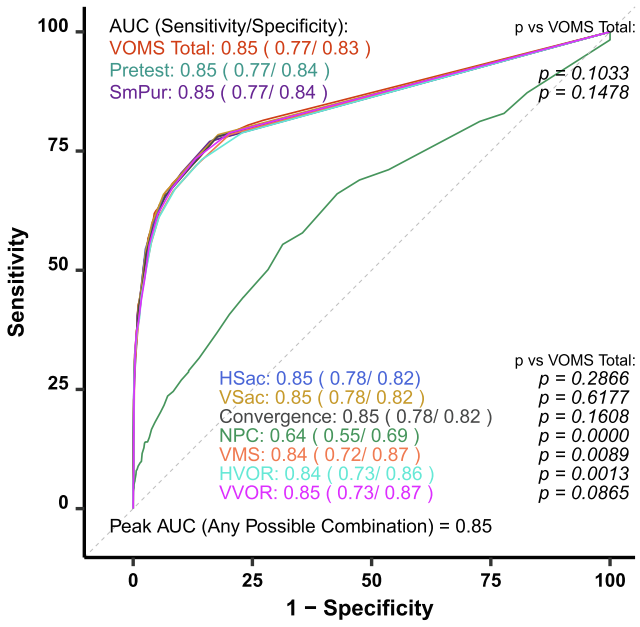

Fig. 5. ROC Analyses for VOMS Items. AUC curves and corresponding sensitivity and specificity for VOMS Total (red), Pretest (teal), Smooth Pursuits (SmPur) (purple), Horizontal Saccades (HSac) (blue), Vertical Saccades (VSac) (yellow), Convergence (Conv) (grey), Mean NPC (NPC) (green), Visual Motion Sensitivity (VMS) (salmon), Horizontal VOR (HVOR) (cyan), and Vertical VOR (VVOR) (magenta). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The results in Aim 1 suggest there is no single, statistics-based approach to reduce the length of VOMS. As suggested by Mucha and colleagues (2014), any combination of four or more oculomotor and vestibular variables performs with near equal diagnostic utility as another. However, when selecting items for a streamlined VOMS tool, ease of use and consistency of testing should be taken into consideration. For example, whereas NPC distance has been suggested as discriminative for concussion identification (Pearce et al., 2015), inconsistencies with measurement, age-related performance variation, and procedural instructions question the quality of information it provides clinicians (Raghuram et al., 2018; Santo et al., 2020). Conversely, easily conducted symptom-based assessments, such as smooth pursuits, horizontal saccades, horizontal VOR and VMS may be a good starting point for keeping the VOMS easy to conduct while retaining high predictive utility.

Another important consideration in streamlining the VOMS is ease of incorporation into existing multi-dimensional sideline assessment batteries such as MACE2 or the SCAT. Results show incorporation of both VOMS and mVOMS into the SCAT3 battery improves the overall diagnostic utility of the combined tool by a similar amount (SCAT3 + VOMS AUC = 0.84, SCAT3 + mVOMS AUC = 0.83). It is important to note these improvements in discriminative utility were maintained even when mVOMS was scaled to a 7-point Likert scale as seen in the SCAT3 Symptom Evaluation. When incorporating an assessment into an existing tool, uniformity of scales should be taken into consideration as it improves ease of use for both clinicians and athletes. Future research should examine re-scaling the VOMS to match other common concussion assessment symptom severity scales.

Whereas cutoff scores for VOMS items have been proposed (Elbin et al., 2021; Kontos, Eagle, et al., 2021; Mucha et al., 2014; National Academies of Sciences, Engineering, and Medicine et al., 2019; Yorke et al., 2017), ambiguity existed as to whether these cutoffs should take pretest symptom severity into consideration. This study is the first to directly compare the diagnostic utility of the tool with and without incorporating pretest symptoms. As demonstrated by the large effect sizes seen with both VOMS Total (d = 2.44) and VOMS Pretest (d = 2.39), assessing headache, dizziness, nausea, and fogginess symptoms alone is highly discriminative for acute concussion. These effect sizes directly correlate to the AUCs seen for both variables (AUC = 0.85, respectively). Symptom severity change scores from pre- to post-exacerbation status yield similar results for both effect size (d = 2.43) and discriminability (AUC = 0.85), suggesting evaluating the VOMS through raw severity scores or through change scores are equally useful for discriminating between acutely concussed and non-concussed populations. Clinicians may want to consider these findings, especially during time-sensitive evaluations where calculating raw sums may be simpler than using change scores.

Although these findings address the ambiguity surrounding a pretest change score, the high effect size and AUC of the Pretest symptoms raise questions as to whether there is any additional benefit during the acute concussion evaluation to conducting the VOMS symptom-exacerbating items at all. Indeed, the overall discriminative utility of both VOMS Total and VOMS Pretest were nearly identical ($p < 0.01$), and as shown in Fig. 10A, VOMS causes only an 11% increase in symptoms post-exacerbation. Yet, when VOMS Pretest and VOMS Total are independently incorporated into the SCAT3 battery (SCAT3 Symptom Evaluation, SAC, mBESS) (Fig. 8), VOMS Pretest (Pretest + SCAT3 AUC = 0.81) yields a lower AUC than either VOMS or mVOMS (SCAT3+mVOMS AUC = 0.83, SCAT3+VOMS AUC = 0.84). This suggests the VOMS exacerbating items provide an additional benefit beyond simply re-assessing symptom severity, albeit in a small percentage of individuals.

In some ways, the VOMS can be considered similar to other

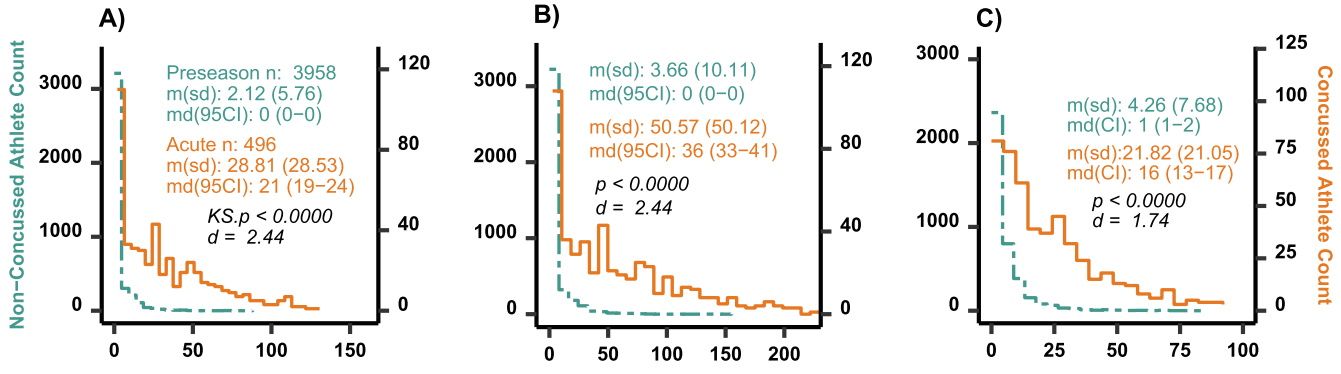

Fig. 6. Distribution and Descriptive Statistics for mVOMS Total, VOMS Total, and SCAT3 Symptom Evaluation. Preseason non-concussed (teal) and acutely concussed (orange) scores for: A) modified VOMS Total (mVOMS Total), B) VOMS Total, and C) SCAT3 Symptom Severity. The left axis displays the count of non-concussed athletes, while the right axis relays the count of concussed athletes. The notation "m" refers to the mean whereas "md" refers to the median. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

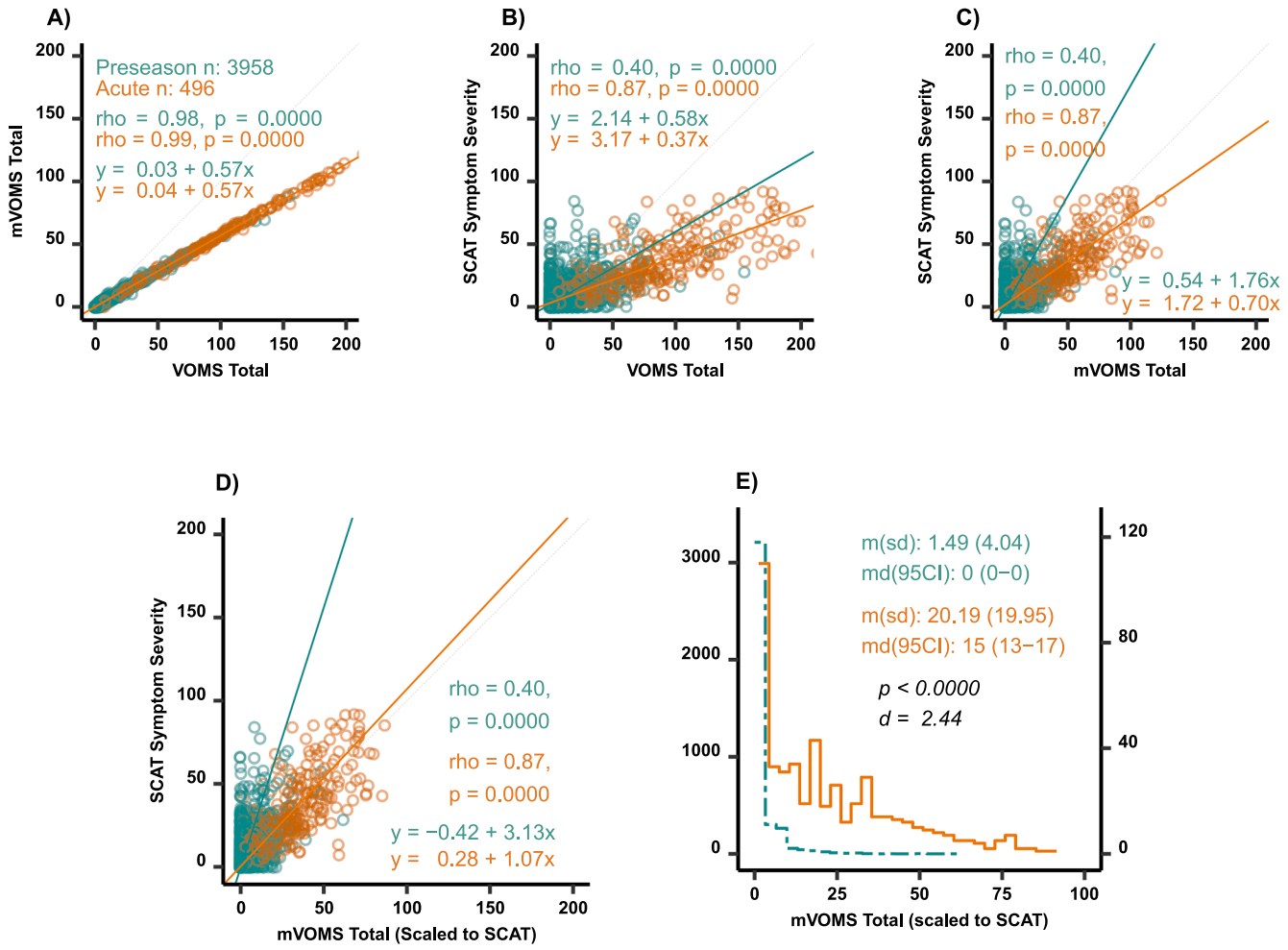

Fig. 7. Deming Regressions, and Distributions and Descriptive Statistics for mVOMS Total, Voms Total and SCAT3 Symptom Evaluation. Preseason non-concussed (teal) and acutely concussed (orange) scores for: A) modified VOMS Total (mVOMS Total) relative to VOMS Total, B) VOMS Total relative to SCAT3 Symptom Severity, C) mVOMS Total and D) a scaled mVOMS Total relative to SCAT3 Symptom Severity. Deming regression formulas and their corresponding lines (solid teal, solid orange) are displayed on each plot in relation to the unity line (light grey). Panel E displays the preseason and acutely concussed scores for the scaled mVOMS Total. The notation "m" refers to the mean whereas "md" refers to the median. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

exacerbating assessment tools, such as the Balke or Buffalo Concussion Treadmill Test (Haider et al., 2019; Leddy et al., 2011), wherein assessing dynamic self-reported symptom scores following intentional engagement in functional movements boosts the sensitivity and specificity of the tools over assessing static symptoms alone. These shifts highlight the sensitive nature of the vestibular and oculomotor systems to stressors and emphasize how VOMS exacerbating tasks are successful both in ruling out non-concussed individuals and in identifying individuals who may require further evaluation for possible underlying conditions such as migraine headache, oculomotor/binocular vision disorders, or vestibular abnormalities (Elbin et al., 2018; Sufrinko et al., 2017). In fact, recent findings suggest the inherent exacerbating nature of VOMS may be amplified if combined with other supervised exercise challenges (Popovich et al., 2021). Together, these findings suggest further use for the VOMS tool beyond identification of vestibular/oculomotor dysfunction in acutely concussed athletes, and beyond discriminating between acutely concussed athletes and healthy controls. Future research should evaluate potential use of the tool during the management phases of concussion recovery.

Integer cutoffs are popular in clinical assessments for their ease of use. This is the first study to directly compare all proposed VOMS cutoff methods utilized in either civilian or military settings (Elbin et al., 2021; Kontos, Eagle, et al., 2021; Mucha et al., 2014; National Academies of Sciences, Engineering, and Medicine et al., 2019; Yorke et al., 2017). It is important to consider the methodology when interpreting ROC results. As shown in Fig. 11A and 12, an AUC generated from raw data is different from one generated from a binary classification process. This is because the raw data method allows for multiple adjustments in cutoff points along the rounded curve, whereas a single cutoff creates a binary process that limits the ROC to a single point on a two-segment line. The similar distributions and effect sizes seen between VOMS Total and mVOMS Total (Fig. 6) foreshadow the mirrored AUCs, sensitivities and specificities seen when comparing potential integer cutoffs between these two measures. With medians and 90th confidence intervals of zero in healthy athletes, any symptom scores beyond 1 for both VOMS Total and mVOMS Total are already discriminative. Increasing cutoffs beyond 1 for both measures varies the sensitivity and specificity of these measures, albeit in small amounts. For VOMS Total, there is no statistically significant difference between choosing clinical integer cutoffs between 4 and 6. Beyond a cutoff of 6, the overall discriminability of VOMS Total decreases due to the drop in sensitivity at the expense of an increased specificity. Integer clinical cutoffs for mVOMS Total follows a similar trend. Clinicians should keep these integer cutoff characteristics for VOMS and mVOMS in mind when assessing performance, especially in situations where it may be desirable to prioritize sensitivity over specificity, or vice versa.

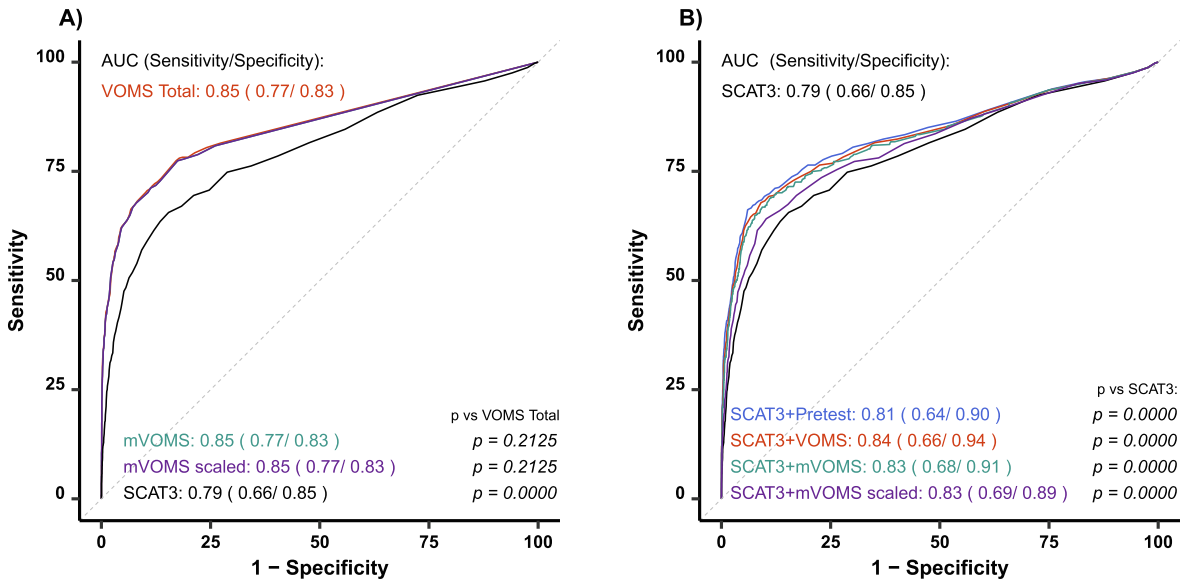

Fig. 8. ROC Analysis for VOMS Tools in relation to the SCAT3 Battery ROC. ROC analyses for A) VOMS Total (orange), modified VOMS (mVOMS) (teal), scaled mVOMS (purple), and the existing SCAT3 battery (black). B) Predictive utilities for variations of the VOMS tool when incorporated into the SCAT3 battery. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

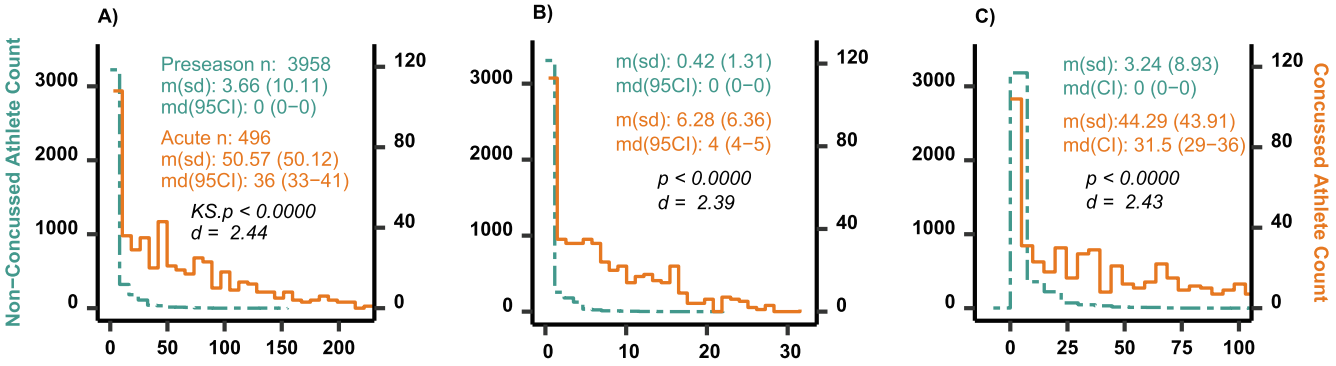

Fig. 9. Distributions and Descriptive Statistics for VOMS Total, VOMS Pretest Total Score, and VOMS Total Change Score. Preseason non-concussed (teal) and acutely concussed (orange) scores for: A) VOMS Total, B) VOMS Pretest Total, and C) VOMS Total Change Score. The left axis displays the count of non-concussed athletes, while the right axis relays the count of concussed athletes. The notation "m" refers to the mean whereas "md" refers to the median. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Finally, there has been a need to compare all proposed VOMS clinical cutoff methods using the same analysis strategy within the same data set to evaluate the discriminability of these methods and determine how they perform in acute concussion evaluations. Directly comparing the utility of different proposed VOMS cutoff methods, including VOMS Total and mVOMS Total cutoffs as identified in Fig. 12, reveals assessing VOMS performance in terms of the sum of symptom severity scores provides higher discriminability for acute concussion than assessing VOMS item by item. This is especially apparent when assessing itemized change scores for VOMS, which yielded low AUCS (Yorke et al., 2017 AUC = 0.69, Elbin et al., 2021 AUC = 0.68) which sustained diagnostic penalties ranging from 5% to 8% compared to the raw itemized methods (Mucha et al., 2014 AUC = 0.76, Kontos et al., 2021 AUC = 0.74, MACE2 AUC = 0.74), and 10–12% compared to any total sum method (AUC ≥ 0.79). Evaluating the three VOMS total sum methods (VOMS Total ≥ 4, mVOMS Total ≥ 4, VOMS Overall ≥ 8) and total sum change method (overall VOMS Change Score ≥ 3) reveals similar overall discriminability (0.79 ≤ AUC ≤ 0.80). That said, slight changes in the respective sensitivities and specificities of these methods should be considered. As previously noted, summed VOMS and mVOMS Total scores are already highly discriminative above symptom reports of 1, with a sensitivity of 81% for both (Fig. 11). A conservative symptom sum of 4 for either of these methods reduces that sensitivity of 78% (VOMS Total) and 77% (mVOMS Total), but improves the specificity (80%, and 83%, respectively) (Fig. 12). This trend continues as cutoffs increase: higher summed VOMS score cutoffs, such as the VOMS Overall cutoff of ≥8, provide higher specificity (86%), at the expense of sensitivity (73%). Clinicians should keep these concepts in mind when determining how to utilize potential VOMS integer cutoffs. In cases where it is desirable to maximize sensitivity, a lower VOMS cutoff score may want to be utilized.

5. Limitations

Several limitations may have impacted this study's analyses and results. First, findings may not be representative of military service members, or of all athletes or age-ranges, as this VOMS investigation is based mostly (~77%) on athletes from NCAA Division I athletic programs. While the findings are encouraging, this analysis should be replicated in a military sample in both deployed/combat and training/garrison settings. Secondly, the CARE study was conducted across multiple sites and years. While all consortium sites and personnel received comprehensive training on protocols and test administration,

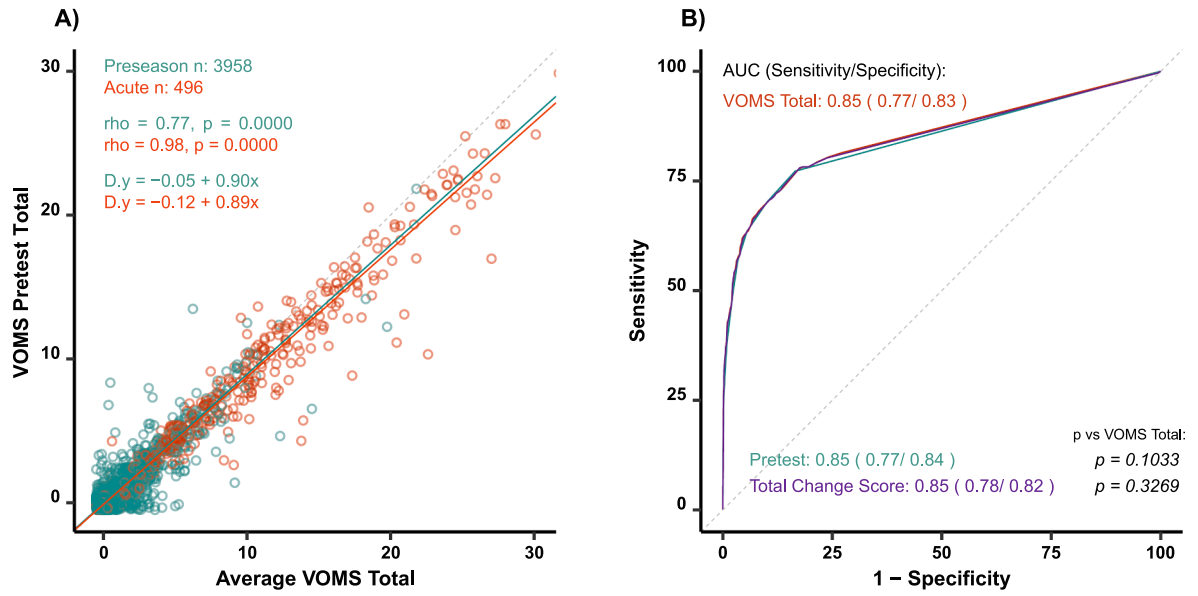

Fig. 10. Deming Regression and ROC Analyses for VOMS Pretest Total, VOMS Total, and VOMS Pretest Change Score. A) Preseason non-concussed (teal) and acutely concussed (orange) symptom severity scores for VOMS Pretest Total and Average VOMS Total. B) AUC curves and corresponding sensitivity and specificity for VOMS Total, VOMS Pretest, and VOMS Pretest Change Score. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

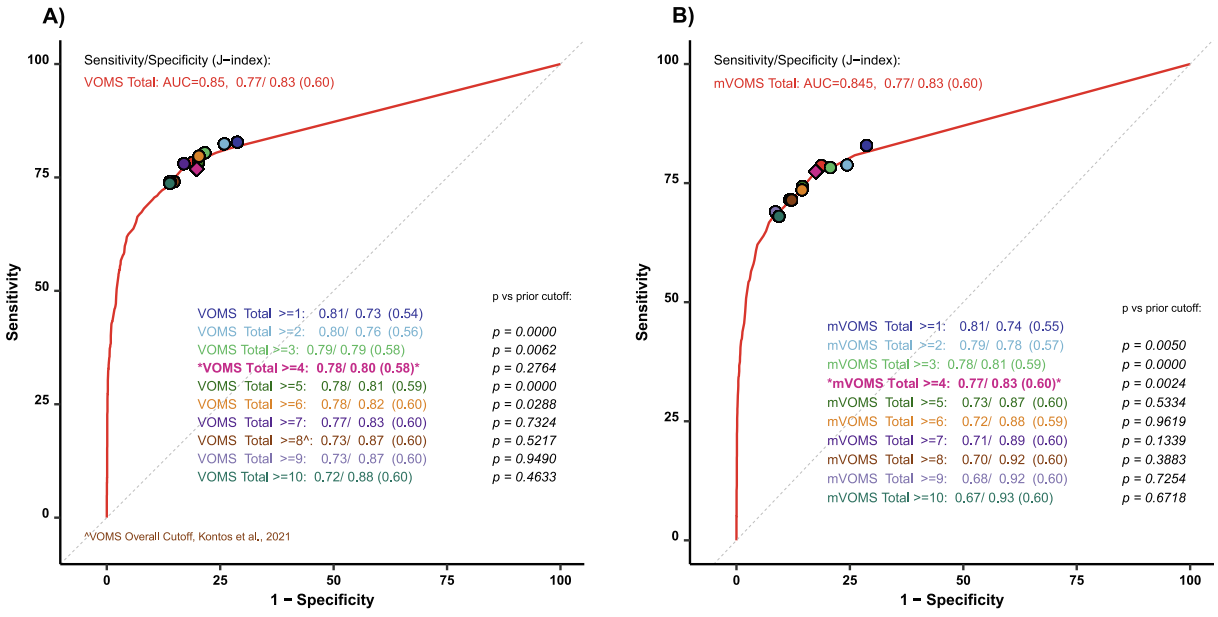

Fig. 11. ROC analyses for VOMS Total relative to mVOMS Total clinical cutoff scores. ROC analyses and significance comparisons for A) Raw VOMS Total (red) relative to integer clinical cutoffs for VOMS Total 1 (blue), 2 (light blue), 3 (light green), 4 (magenta), 5 (dark green), 6 (yellow), 7 (purple), 8 (brown), 9 (light purple), and 10 (teal). B) Raw mVOMS Total (red) relative to integer clinical cutoffs for mVOMS Total 1 (blue), 2 (light blue), 3 (light green), 4 (magenta), 5 (dark green), 6 (yellow), 7 (purple), 8 (brown), 9 (light purple), and 10 (teal). Asterisks note the optimal integer cutoff curve for each as determined by a lack of significant increase in discriminability, respectively. All integer cutoff points are jittered about the raw ROC curve for visibility. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

there is likely variability in test documentation and implementation between sites and administrators. This may explain the lack of results found with NPC distance, for instance. Third, the symptom severity sections in both VOMS and the SCAT3 Symptom Evaluation rely on subjective reports, which are known to be less reliable than results obtained from objective testing. This may have led to individual discrepancies in symptom severity reporting at preseason baseline or acute-SRC timepoints.

Another concern is that all individual VOMS items, and therefore VOMS Total as the symptom severity sum of those items, are non-parametric in their data distribution and show significant skew. Unlike principal component analysis, the regression methods (including the Deming regression utilized in this study) operate with the underlying assumption of normality. This assumption may lead to errors and may partially account for the variability in VOMS item selection when different logistic regression criteria were applied. While this is not ideal, these techniques were incorporated because similar methods have been used in the existing literature, permitting future comparisons.

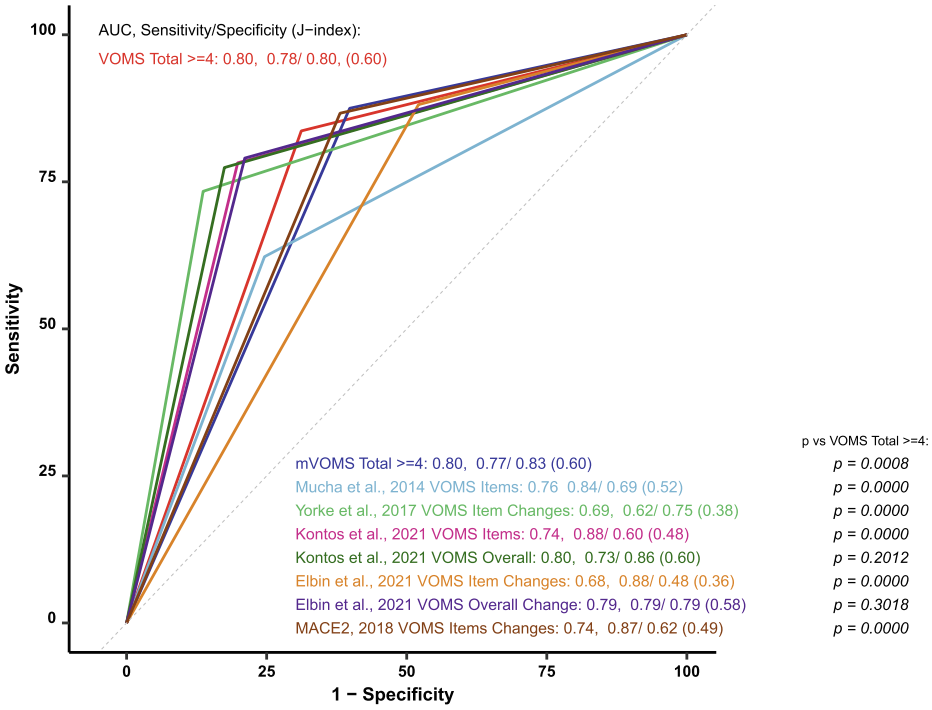

Fig. 12. ROC analyses for proposed VOMS, VOMS Total, and mVOMS Total clinical cutoff scores. integer clinical cutoff two-segment line ROC curves VOMS Total ≥ 4 (red), mVOMS Total ≥ 4 (blue), the Mucha and colleagues individualized VOMS cutoffs (light blue), the Yorke and colleagues individualized VOMS change score cutoffs (light green), the Kontos and colleagues individualized VOMS cutoffs (magenta), the Kontos and colleagues VOMS Overall/Total ≥ 8 cutoff (green), the Elbin and colleagues individualized VOMS change score cutoffs (yellow), the Elbin and colleagues overall VOMS change score (purple), and the MACE2 individualized VOMS cutoffs (brown). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Additionally, as noted in the methods, Deming regression assumes the ratio of measurement error between the two variables being analyzed is constant and is known prior to the investigation. We propose the ratio is similar between VOMS Total and the SCAT3 symptom evaluation given their similar distributions (see Fig. 6), but this assumption could be incorrect.

Finally, the subjective reporting nature of VOMS should be emphasized. This symptom self-report presents a challenge for clinicians evaluating performance. When acute post-concussion symptom burden reports are high, it is easier for clinicians to discriminate between concussed and non-concussed athletes, and it is more difficult for athletes to falsify reports (e.g., falsely report "severe" headache). However, when mild symptoms are reported by athletes, it becomes challenging for clinicians to interpret the legitimacy of the report, even if symptoms are exacerbated by VOMS. This challenge permeates all subjective-report concussion assessments and supports the growing emphasis for the development of reliable and objective assessments for concussion diagnosis and management.

6. Conclusion

This study's results show that VOMS items demonstrate clinically useful discriminative utility for identifying acute concussion. As anticipated, the oculomotor and vestibular VOMS tasks are synergistically linked, giving the tool intractable redundancies. However, a modified VOMS (mVOMS) containing only four items (Smooth Pursuits, Horizontal Saccades, Horizontal VOR, VMS) yields identical AUCs to VOMS Total while shortening administration length by half. In addition, scaling this modified VOMS to match the 0–6 Likert scale used in the SCAT3 Symptom Evaluation maintains this boost in overall predictive utility while unifying symptom scales across the two assessments. Finally, direct comparisons between clinical integer cutoff methodologies revealed use of raw total symptom severity scores for VOMS (≥4) and mVOMS (≥4) to be more discriminative than itemized evaluations of performance, and equally as discriminative as the proposed overall VOMS change score (change ≥ 3).

It is important to note that the VOMS tool was developed to screen patients for vestibular and oculomotor symptoms and impairment following concussion to inform additional assessments and treatment. In contrast, the results of this investigation are limited to the assessment of acute concussion (<72 h) timeframe. Clinicians should keep these differences in mind when considering use of the VOMS tool in athletes beyond the acute phases of injury. As a final point, the VOMS causes an 11% increase in symptoms post-exacerbation. As such, future research into streamlined VOMS models should be evaluated in civilian and military populations at both the acute time frame, as well as in the recovery phases of concussion, to better understand the uses of the tool across the spectrum of concussion. In the interim, clinicians and medics should consider incorporating a modified VOMS into their existing multi-domain acute concussion assessments. Doing so could enable rapid assessments while improving overall concussion identification rates.

Funding

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 Collegiate 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 Combat Casualty Care Research Program, endorsed by the Department of Defense, through the Joint Program Committee 6/ Combat Casualty Care Research Program – Psychological Health and Traumatic Brain Injury Program under Award No. W81XWH1420151. Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the Department of Defense.

CRediT authorship contribution statement

Lyndsey M. Ferris: Conceptualization, Methodology, Data curation, Visualization, Writing – original draft. **Anthony P. Kontos:** Conceptualization, Writing – review & editing. **Shawn R. Eagle:** Writing – review & editing. **R.J. Elbin:** Writing – review & editing. **James R. Clugston:**

Writing – review & editing. **Justus Ortega:** Writing – review & editing. **Nicholas L. Port:** Conceptualization, Data curation, Writing – review & editing.

Acknowledgements

We thank Jaroslaw Harezlak, William H. Swanson, and Thomas A. Buckley for their statistical and conceptual advice.

References

Al Jufali, M., Bishenden, W., Dohi, M., Drezner, J., Patricios, J., Pedrinelli, A., Schmied, C., & Zideman, D. (2015). *Football Emergency Medicine Manual* (2nd Edition). Fédération Internationale de Football Association. https://resources.fifa.com/image/upload/football-emergency-medicine-manual-2nd-edition-2018-2674609.pdf?cloudid=szriio9ble9hvoyns04h.

Bakker, A., Cai, J., English, L., Kaiser, G., Mesa, V., & Van Dooren, W. (2019). Beyond small, medium, or large: Points of consideration when interpreting effect sizes. *Educational Studies in Mathematics, 102*(1), 1–8.

Baldwin, G. T., Breiding, M. J., & Dawn Comstock, R. (2018). Epidemiology of sports concussion in the United States. *Handbook of Clinical Neurology, 158*, 63–74.

Bewick, V., Cheek, L., & Ball, J. (2004). Statistics review 13: Receiver operating characteristic curves. *Critical Care / the Society of Critical Care Medicine, 8*(6), 508–512.

Broglio, S. P., CARE Consortium Investigators, Harezlak, J., Katz, B., Zhao, S., McAllister, T., & McCrea, M. (2019). Acute Sport Concussion Assessment Optimization: A Prospective Assessment from the CARE Consortium. In *Sports Medicine* (Vol. 49, Issue 12, pp. 1977–1987). https://doi.org/10.1007/s40279-019-01155-0.

Broglio, S. P., McCrea, M., McAllister, T., Harezlak, J., Katz, B., Hack, D., Hainline, B., & CARE Consortium Investigators. (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. *Sports Medicine , 47*(7), 1437–1451.

Buckley, T. A., Burdette, G., & Kelly, K. (2015). Concussion-Management Practice Patterns of National Collegiate Athletic Association Division II and III Athletic Trainers: How the Other Half Lives. *Journal of Athletic Training, 50*(8), 879–888.

Caccese, J. B., Eckner, J. T., Franco-MacKendrick, L., Hazzard, J. B., Ni, M., Broglio, S. P., McAllister, T., McCrea, M., Pasquina, P. F., & Buckley, T. A. (2020). Clinical reaction time after concussion: change from baseline versus normative-based Cutoff scores. *Journal of Athletic Training*. https://doi.org/10.4085/JAT0457-20

Cancelliere, C., Coronado, V. G., Taylor, C. A., & Xu, L. (2017). Epidemiology of isolated versus nonisolated mild traumatic brain injury treated in emergency departments in the united states, 2006–2012: sociodemographic characteristics. *The Journal of Head Trauma Rehabilitation, 32*(4), E37–E46.

Capó-Aponte, J. E., Jorgensen-Wagers, K. L., Sosa, J. A., Walsh, D. V., Goodrich, G. L., Temme, L. A., & Riggs, D. W. (2017). Visual Dysfunctions at Different Stages after Blast and Non-blast Mild Traumatic Brain Injury. *Optometry and Vision Science: Official Publication of the American Academy of Optometry, 94*(1), 7–15.

Ciuffreda, K. J., Kapoor, N., Rutner, D., Suchoff, I. B., Han, M. E., & Craig, S. (2007). Occurrence of oculomotor dysfunctions in acquired brain injury: A retrospective analysis. *Optometry, 78*(4), 155–161.

Cochrane, A. (2017). Clinical Procedures for Ocular Examination Nancy B Carlson and Daniel Kurtz McGraw-Hill Education, 4th edn. 2016 581 pages RRP $130.00. In *Clinical and Experimental Optometry* (Vol. 100, Issue 3, pp. 296–296). https://doi.org/10.1111/cxo.12514.

DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. *Biometrics, 44*(3), 837.

Deming, W. E. (1943). *Statistical adjustment of data*. *261*. https://psycnet.apa.org/fulltext/1944-00642-000.pdf.

DoD TBI Worldwide Numbers. (n.d.). Retrieved April 27, 2021, from https://www.health.mil/About-MHS/OASDHA/Defense-Health-Agency/Research-and-Development/Traumatic-Brain-Injury-Center-of-Excellence/DoD-TBI-Worldwide-Numbers.

Echemendia, R. J., Meeuwisse, W., McCrory, P., Davis, G. A., Putukian, M., Leddy, J., Makdissi, M., Sullivan, S. J., Broglio, S. P., Raftery, M., Schneider, K., Kissick, J., McCrea, M., Dvořák, J., Sills, A. K., Aubry, M., Engebretsen, L., Loosemore, M., Fuller, G., … Herring, S. (2017). The Sport Concussion Assessment Tool 5th Edition (SCAT5): Background and rationale. *British Journal of Sports Medicine, 51*(11), 848–850.

Elbin, R. J., Eagle, S. R., Marchetti, G. F., Anderson, M., Schatz, P., Womble, M. N., Stephenson, K., Covassin, T., Collins, M. W., Mucha, A., & Kontos, A. P. (2021). Using change scores on the vestibular ocular motor screening (VOMS) tool to identify concussion in adolescents. *Applied Neuropsychology. Child*, 1–7.

Elbin, R. J., Sufrinko, A., Anderson, M. N., Mohler, S., Schatz, P., Covassin, T., Mucha, A., Collins, M. W., & Kontos, A. P. (2018). Prospective Changes in Vestibular and Ocular Motor Impairment After Concussion. *Journal of Neurologic Physical Therapy: JNPT, 42*(3), 142–148.

Ferris, L. M., Kontos, A. P., Eagle, S. R., Elbin, R. J., Collins, M. W., Mucha, A., Clugston, J. R., & Port, N. L. (2021). Predictive Accuracy of the Sport Concussion Assessment Tool 3 and Vestibular/Ocular-Motor Screening, Individually and In Combination: A National Collegiate Athletic Association-Department of Defense Concussion Assessment, Research and Education Consortium Analysis. *The American Journal of Sports Medicine, 49*(4), 1040–1048.

Ferris, L. M., Kontos, A. P., Eagle, S. R., Elbin, R. J., Collins, M. W., Mucha, A., McAllister, T. W., Broglio, S. P., McCrea, M., Pasquina, P. F., & Port, N. L. (2022 Mar). Utility of VOMS, SCAT3, and ImPACT Baseline Evaluations for Acute Concussion Identification in Collegiate Athletes: Findings From the NCAA-DoD Concussion Assessment, Research and Education (CARE) Consortium. *The American Journal of Sports Medicine, 50*(4), 1106–1119. https://doi.org/10.1177/03635465211072261. Epub 2022 Feb 18 PMID: 35179972.

Gerberich, S. G., Priest, J. D., Boen, J. R., Straub, C. P., & Maxwell, R. E. (1983). Concussion incidences and severity in secondary school varsity football players. *American Journal of Public Health, 73*(12), 1370–1375.

Gottshall, K., Drake, A., Gray, N., McDonald, E., & Hoffer, M. E. (2003). Objective vestibular tests as outcome measures in head injury patients. *The Laryngoscope, 113*(10), 1746–1750.

Greenwald, B. D., Kapoor, N., & Singh, A. D. (2012). Visual impairments in the first year after traumatic brain injury. *Brain Injury: [BI], 26*(11), 1338–1359.

Green, W., Ciuffreda, K. J., Thiagarajan, P., Szymanowicz, D., Ludlam, D. P., & Kapoor, N. (2010). Accommodation in mild traumatic brain injury. *Journal of Rehabilitation Research and Development, 47*(3), 183–199.

Guskiewicz, K. M., Register-Mihalik, J., McCrory, P., McCrea, M., Johnston, K., Makdissi, M., Dvorák, J., Davis, G., & Meeuwisse, W. (2013). Evidence-based approach to revising the SCAT2: Introducing the SCAT3. *British Journal of Sports Medicine, 47*(5), 289–293.

Haider, M. N., Leddy, J. J., Wilber, C. G., Viera, K. B., Bezherano, I., Wilkins, K. J., Miecznikowski, J. C., & Willer, B. S. (2019). The Predictive Capacity of the Buffalo Concussion Treadmill Test After Sport-Related Concussion in Adolescents. *Frontiers in Neurology, 10*, 395.

Harmon, K. G., Clugston, J. R., Dec, K., Hainline, B., Herring, S. A., Kane, S., Kontos, A. P., Leddy, J. J., McCrea, M. A., Poddar, S. K., Putukian, M., Wilson, J. C., & Roberts, W. O. (2019). American Medical Society for Sports Medicine Position Statement on Concussion in Sport. *Clinical Journal of Sport Medicine: Official Journal of the Canadian Academy of Sport Medicine, 29*(2), 87–100.

Heitger, M. H., Anderson, T. J., Jones, R. D., Dalrymple-Alford, J. C., Frampton, C. M., & Ardagh, M. W. (2004). Eye movement and visuomotor arm movement deficits following mild closed head injury. *Brain: A Journal of Neurology, 127*(Pt 3), 575–590.

Ialongo, C. (2016). Understanding the effect size and its measures. *Biochemia Medica: Casopis Hrvatskoga Drustva Medicinskih Biokemicara / HDMB, 26*(2), 150–163.

Kassambara, A. (2017). Practical Guide To Principal Component Methods in R: PCA, M (CA), FAMD, MFA, HCPC, factoextra. STHDA.

Kelly, K. C., Jordan, E. M., Joyner, A. B., Burdette, G. T., & Buckley, T. A. (2014). National Collegiate Athletic Association Division I athletic trainers' concussion-management practice patterns. *Journal of Athletic Training, 49*(5), 665–673.

Kelly, K. M., Kiderman, A., Akhavan, S., Quigley, M. R., Snell, E. D., Happ, E., Synowiec, A. S., Miller, E. R., Bauer, M. A., Oakes, L. P., Eydelman, Y., Gallagher, C. W., Dinehart, T., Schroeder, J. H., & Ashmore, R. C. (2019). Oculomotor, Vestibular, and Reaction Time Effects of Sports-Related Concussion: Video-Oculography in Assessing Sports-Related Concussion. *The Journal of Head Trauma Rehabilitation, 34*(3), 176–188.

Kontos, A. P., Eagle, S. R., Marchetti, G., Sinnott, A., Mucha, A., Port, N. L., Ferris, L. M., Elbin, R. J., Clugston, J., Ortega, J., Broglio, S. P., McAllister, T., McCrea, M., Pasquina, P., Collins, M. W., CARE Consortium Investigators:, Buckley, T., Mihalik, J., & Miles, C. (2021). Discriminative utility of VOMS in identifying concussed from healthy collegiate athlete controls: an NCAA-DoD CARE Consortium study. *The American Journal of Sports Medicine, EPub ahead of print*. https://doi.org/10.1177/03635465211012359.

Kontos, A. P., Monti, K., Eagle, S. R., Thomasma, E., Holland, C. L., Thomas, D., Bitzer, H. B., Mucha, A., & Collins, M. W. (2021). Test-retest reliability of the Vestibular Ocular Motor Screening (VOMS) tool and modified Balance Error Scoring System (mBESS) in US military personnel. *Journal of Science and Medicine in Sport / Sports Medicine Australia, 24*(3), 264–268.

Kontos, A. P., Sufrinko, A., Elbin, R. J., Puskar, A., & Collins, M. W. (2016). Reliability and Associated Risk Factors for Performance on the Vestibular/Ocular Motor Screening (VOMS) Tool in Healthy Collegiate Athletes. *The American Journal of Sports Medicine, 44*(6), 1400–1406.

Langlois, J. A., Rutland-Brown, W., & Wald, M. M. (2006). The epidemiology and impact of traumatic brain injury: A brief overview. *The Journal of Head Trauma Rehabilitation, 21*(5), 375–378.

Leddy, J. J., Baker, J. G., Kozlowski, K., Bisson, L., & Willer, B. (2011). Reliability of a graded exercise test for assessing recovery from concussion. *Clinical Journal of Sport Medicine: Official Journal of the Canadian Academy of Sport Medicine, 21*(2), 89–94.

Leigh, J. R., & Zee, D. S. (2015). *The Neurology of Eye Movements*. Oxford University Press.

Lepore, F. E. (1995). Disorders of ocular motility following head trauma. *Archives of Neurology, 52*(9), 924–926.

Lê, S., Josse, J., Husson, F., et al. (2008). FactoMineR: An R package for multivariate analysis. *Journal of Statistical Software, 25*(1), 1–18.

Linnet, K. (1993). Evaluation of regression procedures for methods comparison studies. *Clinical Chemistry, 39*(3), 424–432.

Mandrekar, J. N. (2010). Receiver operating characteristic curve in diagnostic test assessment. *Journal of Thoracic Oncology: Official Publication of the International Association for the Study of Lung Cancer, 5*(9), 1315–1316.

Marín-Franch, I., Malik, R., Crabb, D. P., & Swanson, W. H. (2013). Choice of statistical method influences apparent association between structure and function in glaucoma. *Investigative Ophthalmology & Visual Science, 54*(6), 4189–4196.

Massof, R. W., & Emmel, T. C. (1987). Criterion-free parameter-free distribution-independent index of diagnostic test performance. *Applied Optics, 26*(8), 1395. https://doi.org/10.1364/ao.26.001395

McCrory, P., Meeuwisse, W., Dvořák, J., Aubry, M., Bailes, J., Broglio, S., … Vos, P. E. (2017). Consensus statement on concussion in sport-the 5 international conference on concussion in sport held in Berlin, October 2016. *British Journal of Sports Medicine, 51*(11), 838–847.

McDevitt, J., Appiah-Kubi, K. O., Tierney, R., & Wright, W. G. (2016). Vestibular and Oculomotor Assessments May Increase Accuracy of Subacute Concussion Assessment. *International Journal of Sports Medicine, 37*(9), 738–747.

Merezhinskaya, N., Mallia, R. K., Park, D., Bryden, D. W., Mathur, K., & Barker 2nd, F. M. (2019). Visual Deficits and Dysfunctions Associated with Traumatic Brain Injury: A Systematic Review and Meta-analysis. *Optometry and Vision Science: Official Publication of the American Academy of Optometry, 96*(8), 542–555.

Moran, R. N., Covassin, T., Elbin, R. J., Gould, D., & Nogle, S. (2018). Reliability and Normative Reference Values for the Vestibular/Ocular Motor Screening (VOMS) Tool in Youth Athletes. *The American Journal of Sports Medicine, 46*(6), 1475–1480.

Mucha, A., Collins, M. W., Elbin, R. J., Furman, J. M., Troutman-Enseki, C., DeWolf, R. M., Marchetti, G., & Kontos, A. P. (2014). A Brief Vestibular/Ocular Motor Screening (VOMS) assessment to evaluate concussions: Preliminary findings. *The American Journal of Sports Medicine, 42*(10), 2479–2486.

Murray, N. G., Szekely, B., Islas, A., Munkasy, B., Gore, R., Berryhill, M., & Reed-Jones, R. J. (2020). Smooth Pursuit and Saccades after Sport-Related Concussion. *Journal of Neurotrauma, 37*(2), 340–346.

National Academies of Sciences, Engineering, and Medicine, Health and Medicine Division, Board on Health Care Services, & Committee on the Review of the Department of Veterans Affairs Examinations for Traumatic Brain Injury. (2019). *Evaluation of the Disability Determination Process for Traumatic Brain Injury in Veterans*. National Academies Press.

National Concussion Surveillance System. (2019, March 7). https://www.cdc.gov/traumaticbraininjury/ncss/index.html.

Paluszek, M., & Thomas, S. (2020). *MATLAB Machine Learning Toolboxes.*. https://doi.org/10.1007/978-1-4842-5124-9_2

Pearce, K. L., Sufrinko, A., Lau, B. C., Henry, L., Collins, M. W., & Kontos, A. P. (2015). Near Point of Convergence After a Sport-Related Concussion: Measurement Reliability and Relationship to Neurocognitive Impairment and Symptoms. *The American Journal of Sports Medicine, 43*(12), 3055–3061.

Perkins, N. (2004). The Youden Index and the Corresponding Optimal Cut-Point Corrected for Measurement Error. *Epidemiology, 15*(4), S153.

Popovich, M., Almeida, A., Lorincz, M., Eckner, J. T., Freeman, J., Streicher, N., & Alsalaheen, B. (2021). Does Exercise Increase Vestibular and Ocular Motor Symptom Detection After Sport-Related Concussion? *Journal of Neurologic Physical Therapy: JNPT.*. https://doi.org/10.1097/NPT.0000000000000356

Raghuram, A., Cotter, S. A., & Shah, A. S. (2018). Is a Receded Near Point of Convergence (NPC) Diagnostic of Convergence Insufficiency in Post-Concussion Patients? *Investigative Ophthalmology & Visual Science, 59*(9), 1953.

R Core Team. (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/.

Romme, A. G. L. (1995). Boolean comparative analysis of qualitative data. *Quality and Quantity, 29*(3), 317–329.

Santo, A. L., Race, M. L., & Teel, E. F. (2020). Near Point of Convergence Deficits and Treatment Following Concussion: A Systematic Review. *Journal of Sport Rehabilitation, 29*(8), 1179–1193.

Sawilowsky, S. S. (2009). New effect size rules of thumb. *Journal of Modern Applied Statistical Methods: JMASM, 8*(2), 597–599.

StepReg: Stepwise Regression Analysis. (n.d.). Retrieved May 12, 2021, from https://cran.r-project.org/web/packages/StepReg/index.html.

Sufrinko, A. M., Mucha, A., Covassin, T., Marchetti, G., Elbin, R. J., Collins, M. W., & Kontos, A. P. (2017). Sex Differences in Vestibular/Ocular and Neurocognitive Outcomes After Sport-Related Concussion. *Clinical Journal of Sport Medicine: Official Journal of the Canadian Academy of Sport Medicine, 27*(2), 133–138.

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S, Fourth Edi*. Springer, New York.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer.

Yorke, A. M., Smith, L., Babcock, M., & Alsalaheen, B. (2017). Validity and Reliability of the Vestibular/Ocular Motor Screening and Associations With Common Concussion Screening Tools. *Sports Health, 9*(2), 174–180.