ABSTRACT

Objective The Buffalo Concussion Physical Examination (BCPE) is a brief, but pertinent physical examination designed for the subacute, outpatient assessment of concussion. The purpose of this study was to perform the BCPE on a larger sample and derive a scoring system to identify children at risk for Persistent Post-Concussive Symptoms (PPCS, recovery >30 days).

Methods This prospective, observational cohort study from September 2016 to March 2019 was performed at three university-affiliated concussion clinics. Male and female children (n=270, 14.92±1.86 years, range 8–18, 38% female) were diagnosed with a concussion within 14 days of injury and followed-up until recovery. Logistic regression was used with history and physical examination variables to predict PPCS and a weighted scoring metric was derived.

Results Out of 15 predictor variables, the main effects of 1 preinjury variable (≥3 previous concussions), 2 injury characteristic variables (days-since-injury and type-of-injury), 3 physical examination variables (orthostatic intolerance (OI), vestibulo-ocular reflex (VOR) and tandem gait) and 2 interaction terms (OI/VOR and tandem gait/type-of-injury) produced a score that was 85% accurate for identifying children with low-risk, medium-risk and high-risk for PPCS on cross-validation.

Conclusion The Risk for Delayed Recovery (RDR)-Score allows physicians in an outpatient setting to more accurately predict which children are at greater risk for PPCS early after their injury, and who would benefit most from targeted therapies. The RDR-Score is intended to be used as part of a comprehensive assessment that should include validated symptom checklists, mental health history and adjunct testing (eg, cognitive or physical exertion) where clinically indicated.

INTRODUCTION

Concussion, a form of traumatic brain injury, is a public health concern. It is estimated that approximately 5%–10% of children in the USA will experience a concussion in their lifetime. Children have the highest rates of concussion in sport and tend to require longer to recover than adults. The typical duration of recovery in children is less than 1 month but approximately 30% take longer to recover; this is called Persistent Post-Concussive Symptoms (PPCS). Children with PPCS are far more likely to experience psychosocial adjustment issues and learning difficulties in school. There are currently no objective blood or imaging biomarkers to diagnose concussion or to identify patients early who will take longer to recover.

Concussion diagnosis is a clinical determination based on a thorough history, symptom checklists and behavioural screens, a physical examination and adjunct testing if applicable. Concussion management that includes early and effective planning for school, social and/or sport adjustment for those with delayed recovery is known to effectively reduce stress for the child and the family and improve outcome. Components of what we call PPCS planning are individualised for each patient based on their predominant impairments. They include: (1) monitoring symptom precipitators, such as excessive screen use; (2) sleep hygiene strategies; (3) communication with school personnel regarding specific accommodations; (4) interventions for specific oculomotor, vestibular and/or exercise intolerance deficits and (5) counselling for improving social well-being. Since physicians who see concussions infrequently may not be familiar with recommended rehabilitation strategies, an important aspect of PPCS planning includes referral to specialists (eg, vision and vestibular therapists) with experience in concussion management.

Zemek et al derived the SP risk assessment score consisting of nine variables (age, sex, two medical history components, four self-reported symptoms and tandem stance) used within 48-hours of injury that was 68% accurate predicting PPCS. This tool has also been shown to be effective in the subacute outpatient setting with 75% accuracy. The Buffalo Concussion Physical Examination (BCPE) is a brief (5–7 min), pertinent physical examination protocol designed for the outpatient setting. It assesses subsystems commonly affected after concussion head injury (eg, oculomotor and vestibular) and also includes key red flags for more serious injury that may prompt urgent referral. The BCPE was derived and validated on a cohort of adolescents with sport-related concussion (SRC). The purpose of this study was to perform the BCPE on a larger, more diverse sample of children to derive a scoring metric, independent of self-reported symptoms that within 14 days of injury could identify patients who would benefit from early PPCS planning. We hypothesised that this weighted scoring system, termed the Risk of Delayed Recovery Score
activity12 and to do social/work activities below their symptom-
further head injury, to engage in light aerobic exercise/physical
not participate in sports or other activities that pose a risk for
among all study physicians. Participants were instructed to
assessment,17 and adjunct tests if indicated (eg, exercise toler-
ance test).19 Clinical management protocols were standardised
release if they (1) had a second concussive head
recovery from concussion. Participants were considered recov-
horizontal repetitive saccades; (4) vertical repetitive saccades;
vestibulo-ocular reflex (VOR); (6) near point convergence;
(7) tandem gait;21 (8) neck tenderness; (9) neck muscle spasm
(10) neck range of motion. Directions for each component of
the specific physical examination components that made it to
the final scoring tool are provided in online supplemental file 2.

Main outcome
The main outcome measure was duration of recovery, calculated
as the difference in days from injury to determination of clinical
recovery from concussion. Participants were considered recov-
cerally by their physician when they: (1) returned to a
baseline level of symptoms at rest; (2) had a normal physical
examination and (3) were able to exercise and return-to-school
without exacerbation of concussion-like symptoms.25 Athlete
participants who needed clearance to begin a return-to-play
protocol also had to demonstrate good exercise tolerance.19
Since duration of recovery is subject to interval censoring in
patients with longer recovery times (ie, not knowing the exact
date of recovery due to large gaps between clinic visits after the
first 4 weeks), the outcome used for predictive models was the
binary variable PPCS (recovery time ≥30 days). This is elabo-
rated on in online supplemental file 1 (pp 2–3).

Statistical analysis
Sample size was estimated based on number of components in
the initial assessment for logistic regression using the 10 events
per predictor variable per group rule of thumb28 29 (4 demo-
graphic and 9 physical examinations) for a minimum sample
size of 130 participants in each group (total 260 in two groups).
This a priori sample size estimation did not account for unequal
sample sizes, which is a major limitation of this manuscript and
is discussed in more detail in the limitations section. Physical
examination and demographic variables for the entire popu-
lation were summarised using descriptive statistics. Recovery
times were categorised as normal recovery (≤29 days) and

Demographics
Variables included: (1) age, (2) sex (male/female) and (3) number
of previous concussions.

Injury characteristics
This included time-since-injury and type-of-injury. Time-since-
injury is the difference in days from injury to initial clinic visit.
Type-of-injury was classified according to injury severity into
two groups: (1) single/low-velocity injury and (2) multiple/
high-velocity injury. The latter included complex injuries such as
multiple suspected concussive events without time for recovery
in between (ie, overlapping concussion syndrome)22 and high-
velocity injuries such as motor vehicle accidents (MVAs) or
falls from a height. In the context of SRC, most were classi-
ified as single/low-velocity except if the injury was very serious
(described in detail in the limitations section) or if children
who continued to play after the initial concussive head injury
suffered another head injury within the same game. Children
who continued to play without sustaining another distinct head
injury were classified as single/low-velocity.

Physical examination
Predictor variables from the BCPE included (1) orthostatic intol-
erance23 (OI, defined as reporting symptoms of lightheadedness
or dizziness24 using the 1 min standing25 after 2 min supine26
orthostatic manoeuvre); (2) ocular smooth pursuits (SP); (3)
horizontal repetitive saccades; (4) vertical repetitive saccades;
(5) vestibulo-ocular reflex (VOR); (6) near point convergence;
(7) tandem gait;21 (8) neck tenderness; (9) neck muscle spasm
and (10) neck range of motion. Directions for each component of
the BCPE have been published previously17 and directions for
the specific physical examination components that made it to
the final scoring tool are provided in online supplemental file 2.

Exposure and management of concussion
The exposure of interest for this study was a concussion diag-
nosed using recent Concussion in Sport Group guidelines,7 8
including history (using a symptom checklist and behavioural
screening questionnaires), concussion-like symptoms linked to
a concussive head injury or injury to another part of the body
with force transmitted to the head, a concussion-focused clinical
assessment,7 17 and adjunct tests if indicated (eg, exercise toler-
ance test).17 Clinical management protocols were standardised
among all study physicians. Participants were instructed to
not participate in sports or other activities that pose a risk for
further head injury, to engage in light aerobic exercise/physical
activity17 and to do social/work activities below their symptom-
exacerbation thresholds. Following good medical standards of
practice, cervical impairments (if present) were treated early with
pharmacological and/or non-pharmacological therapies since
untreated cervical injuries are a common comorbidity associated
with delayed recovery.20 Details regarding our practice’s cervical
management protocol are described in online supplemental
file 2. If participants did not recover by 4 weeks from injury, a
multidisciplinary approach was used that included, when cli-
cally indicated, a physical therapist (for vestibular and cervical
therapy), an occupational therapist (for vision therapy) and/or
a neuropsychologist (for cognitive or mood-related issues).21 The
lead study physician (JJL) met with study physicians at quar-
terly department meetings to review/revise clinical management
protocols per university teaching standards.

Predictor variables
Predictor variables were separated into three categories: (1)
demographics; (2) injury characteristics; and (3) physical exam-
ination findings.
delayed recovery (PPCS, ≥30 days). Distribution of the continuous outcome variable (recovery time in days) was assessed, and appropriate transformations were used if applicable. Results of the assessment of distribution and transformations are presented in online supplemental file 1 (pp 2–3).

Individual predictors were regressed to response variables. The distribution of discrete variables (time-since-injury and number of previous concussions) was regressed on response variables and acceptable cut-offs/stratifications were selected to maximise predictive potential. Time-since-injury was not a covariate in the same sense as the other predictors and was regressed on the response variables using Cox Proportional Hazard (Cox PH), Accelerated Failure Time (AFT) and binomial Generalised Linear Modelling (bGLM) models. The Cox PH and AFT models were considered as they are commonly used to model time to event data, while the bGLMs provided greater flexibility than logistic regression alone. For Cox PH, estimated hazard of stratified time-since-injury was plotted against recovery time and PH assumptions were assessed using Schoenfeld Global Test. For AFT and bGLM models, time-since-injury was incorporated as a continuous predictor and the linearity assumption was checked visually using a scatter plot of log recovery time against time-since-injury with a locally estimated scatterplot smoothing curve fitted on top. Four different AFT models were assessed; log-logistic, lognormal, Weibull and Inverse-Weibull which correspond to logistic, normal, extreme value (minimum) and extreme value (maximum) residual distributions, respectively, when fitting a linear regression model to the log-transformed recovery times. For the bGLM model, four different link functions were tested; logit, probit, log-log and complementary log-log that correspond to the survival functions for the log-logistic, lognormal, Weibull and Inverse-Weibull AFT models, respectively. Since the physical examination results of the BCPE can be expected to vary depending on how long after the injury the initial examination is performed, tests for interaction were performed between time-since-injury and these response variables to justify their inclusion as additive terms in the models.

Predictive models were built using Cox PH, AFT and bGLM for incidence of delayed recovery using: (1) physical examination components only; (2) all predictors and (3) all predictors and interactions. Forward stepwise selection was performed using the Akaike Information Criterion (AIC) to identify significant contributors to the model and the best model was selected to build a scoring system. Model fit was assessed visually for AFT models using quantile-quantile plots and for bGLM using the Hosmer-Lemeshow test. Collinearity was assessed by testing for independence between model predictors. Coefficients were multiplied by a common multiplier m to get integer values and RDR-score ranges for low risk (<30%), medium risk (30%–70%) and high risk (>70%). Finally, cross-validation was performed using leave-one-out cross-validation to estimate final model performance and discrimination assessed using the C statistic. Detailed results of the statistical analysis are provided in online supplemental file 1 and a brief summary is provided below. All data analyses were performed using the R programming language.

RESULTS

From September 2016 to March 2019, 359 children and adolescents were evaluated in our university clinics and diagnosed with a concussion. Two hundred and eighty-four participants met eligibility criteria and were included in the study; however, 11 participants were lost to follow-up and three participants had a second head injury while they were still recovering and were removed from analysis. Hence, 270 concussed children and adolescents within 14 days of their injury (majority SRC) comprised the final sample. The 14 participants not included in the analysis did not significantly differ in age or sex. Sample demographics are presented in table 1.

Results of regression of individual predictors on response variables are presented in online supplemental file 1. A cut-off of or more previous concussions was identified to have the highest accuracy. P-values of the tests of independence between predictor and response variables are presented in figure 1. OI, sex, SPs, type of injury and time-since-injury all had a statistically significant association with log of recovery times at a Bonferroni-corrected significance level. OI, SP, type-of-injury, time-since-injury and VOR had a statistically significant association with incidence of delayed recovery at a Bonferroni-corrected significance level.

Physical examination covariates did not reach significance to be time-dependent; hence, time-since-injury was incorporated into the model as an independent covariate. Correlation between predictors was assessed and only two pairs of interaction (sex/neck tenderness and OI/VOR; online supplemental file 1, figure 5.3) were significant at a p<0.05 level. Results of model selection predicting delayed recovery are presented in table 2.

Delayed recovery was predicted for the Cox and AFT models if the median recovery time under the model was ≥30 days, and if the probability was >0.5 for bGLM. The best model was chosen based on highest accuracy and lowest AIC, which was the bGLM model with a complementary log-log link function using all predictor variables (history and physical examination) and interactions with stepwise selection. The C-statistic for this model was 0.74. A calibration plot for this model is presented in figure 2.

The best model predicting delayed recovery included the main effects of one demographic variable (≥3 previous concussion), two injury characteristic variables (time-since-injury and type of injury), three physical examination variables (OI, VOR and tandem gait)
and two interaction terms. Details of the bGLM model with a complementary log–log link function are presented in table 3.

Coefficients from the bGLM model were multiplied by multipliers $m$ to produce scaled, integer coefficients. An $m = 3.832$ was chosen because it was the smallest multiplier with a low root mean square difference between the scaled coefficients and the nearest integers. Scaled integers were tabulated and ranges for low-risk, medium-risk and high-risk were made. Final values for the BCPE screening tool predicting delayed recovery are presented in table 4. After internal cross-validation, the BCPE RDR-Score for delayed recovery was 85% accurate. It misclassified 5% of participants as high-risk for delayed recovery (95% specificity) and misclassified 33% of participants as low-risk for delayed recovery (67% sensitivity).

**DISCUSSION**

This prospective cohort study identified three variables from the history and three variables from a brief physical examination performed a mean of 6 days after injury that predicted PCS with 85% accuracy. Current clinical practice recommendations include delaying focused therapies such as vestibular therapy or cognitive rehabilitation until 1 month after injury. Therefore, experienced clinicians felt that there would be less harm in delaying treatment for a few weeks than prescribing therapies that would not impact the natural recovery seen in most youth patients after concussion. The high specificity of the RDR-score helps achieve this aim by minimising false positive assessment of high risk. We used the cut-off points that maximised accuracy of the decision rule. The ultimate decision regarding the provision of resources and services is up to the clinician and should be individualised to the needs of the patient. The purpose of this study was to give clinicians key information within the subacute evaluation phase after concussion to make treatment decisions to help patients at greater risk for PCS avoid delayed recovery.

![Figure 1](http://bjsm.bmj.com/)

**Figure 1** P-values of the tests of independence between all predictor variables and response variables with a Bonferroni-corrected level. P-values are plotted on a log scale, so the longer bars correspond to smaller p-values.

---

**Table 2** Summary of model selection

<table>
<thead>
<tr>
<th>Model</th>
<th>Binomial GLM (complementary log-log)</th>
<th>AFT model (Inverse-Weibull)</th>
<th>Cox PH model (stratified by time-since-injury)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accuracy (%)</td>
<td>Sensitivity (%)</td>
<td>Specificity (%)</td>
</tr>
<tr>
<td>Null model</td>
<td>64</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>Physical examination only</td>
<td>67 (63)</td>
<td>35 (31)</td>
<td>86 (82)</td>
</tr>
<tr>
<td>All predictors</td>
<td>76 (73)</td>
<td>54 (49)</td>
<td>89 (87)</td>
</tr>
<tr>
<td>Models including interaction terms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All interactions</td>
<td>92 (66)</td>
<td>87 (40)</td>
<td>96 (77)</td>
</tr>
<tr>
<td>Stepwise selected models</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All predictors stepwise</td>
<td>77 (75)</td>
<td>57 (53)</td>
<td>89 (86)</td>
</tr>
<tr>
<td>All interactions stepwise</td>
<td>78 (75)</td>
<td>56 (51)</td>
<td>91 (89)</td>
</tr>
</tbody>
</table>

*Bolded values indicate the final model used for the RDR score.*

AFT, accelerated failure time; AIC, Akaike information criterion; GLM, generalised linear modelling; PH, proportional hazard.

---

so the lower bound on the accuracy of any predictive model was set at 64%, simply by predicting that no participants would have delayed recovery regardless of their risk factors. Our BCPE RDR-score was 85% accurate when calculated during the early phase after injury after cross-validation, which was significantly better than the null hypothesis for predicting PPCS incidence.

PPCS planning should be based on the patient’s underlying impairments. This convenient risk calculation score identifies children who require early PPCS planning, but it can also be used with other clinical tools, including symptom checklists that identify mood-related and cognition-related impairments, mental health and migraine history and treadmill testing that identifies exercise intolerance. The information is used to individualise a treatment programme to reduce the risk of prolonged recovery. Children identified by the RDR-score as low-risk for PPCS should nevertheless be given low-resource guidance to help ensure that they will recover uneventfully, including education about avoiding symptom provocation, safe physical and cognitive activity levels, basic sleep hygiene and how to monitor recovery. A mobile calculator app (eg, MD Calc) or an Excel spreadsheet could be used to make a template with ‘Yes or No’ responses to the six items (with appropriate multipliers) to rapidly provide the RDR-score for physicians in the busy clinic setting. Two sample clinical scenarios using the BCPE RDR-score are provided in online supplemental file 2).

Some variables not included in our model have previously been identified as predictors of PPCS. Female sex was a significant contributor to predicting PPCS in the Cox PH and AFT models, but not in the bGLM model, so it did not make it into the RDR-score. Female sex has historically been associated with longer recovery times and it correlated with longer mean recovery time in our sample, but it was not predictive of the binary diagnosis of PPCS (≥30 days recovery). In a recent prospective study, however, male and female patients managed with early aerobic exercise recovered equally well. Participants in our study were managed according to our clinics’ protocols, which identify and treat symptom generators regardless of sex. There was also a

### Table 3 Details of the final bGLM model with a complementary log-log link function

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coeff.</th>
<th>P &gt;</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time-since-injury</td>
<td>0.26</td>
<td>&lt;0.001</td>
<td>0.19 to 0.33</td>
</tr>
<tr>
<td>OI</td>
<td>1.31</td>
<td>&lt;0.001</td>
<td>0.58 to 2.03</td>
</tr>
<tr>
<td>VOR</td>
<td>1.29</td>
<td>&lt;0.001</td>
<td>0.56 to 2.02</td>
</tr>
<tr>
<td>≥3 previous concussions</td>
<td>1.08</td>
<td>0.018</td>
<td>0.18 to 2.0</td>
</tr>
<tr>
<td>Tandem gait</td>
<td>0.29</td>
<td>0.259</td>
<td>−0.22 to 0.8</td>
</tr>
<tr>
<td>High-velocity/multiple impact</td>
<td>0.50</td>
<td>0.023</td>
<td>−0.43 to 1.43</td>
</tr>
<tr>
<td>OI × VOR</td>
<td>−1.13</td>
<td>0.023</td>
<td>−2.11 to −0.17</td>
</tr>
<tr>
<td>High-velocity/multiple impact × tandem gait</td>
<td>1.27</td>
<td>0.103</td>
<td>−0.25 to 2.79</td>
</tr>
<tr>
<td>Intercept</td>
<td>−3.69</td>
<td>&lt;0.001</td>
<td>−4.50 to −2.89</td>
</tr>
</tbody>
</table>

OI, orthostatic intolerance; VOR, vestibulo-ocular reflex.

### Table 4 The Buffalo Concussion Physical Examination RDR-score for delayed recovery

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Description</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Days since injury</td>
<td>n (in days)</td>
<td>1 point per day</td>
</tr>
<tr>
<td>High velocity/multiple impact</td>
<td>No (0) or yes (1)</td>
<td>2 points</td>
</tr>
<tr>
<td>≥3 previous concussions</td>
<td>No (0) or yes (1)</td>
<td>4 points</td>
</tr>
<tr>
<td>OI</td>
<td>Normal (0) or abnormal (1)</td>
<td>5 points</td>
</tr>
<tr>
<td>VOR</td>
<td>Normal (0) or abnormal (1)</td>
<td>5 points</td>
</tr>
<tr>
<td>Tandem gait</td>
<td>Normal (0) or abnormal (1)</td>
<td>1 point</td>
</tr>
<tr>
<td>OI × VOR</td>
<td>1 if both are abnormal, 0 otherwise</td>
<td>−4 points</td>
</tr>
<tr>
<td>High velocity/multiple impact × tandem gait</td>
<td>1 if both are abnormal, 0 otherwise</td>
<td>5 points</td>
</tr>
</tbody>
</table>

RDR-score classifications

<table>
<thead>
<tr>
<th>Total score range</th>
<th>OI, orthostatic intolerance; RDR, risk for delayed recovery; VOR, vestibulo-ocular reflex.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low risk (&lt;30% risk of delayed recovery)</td>
<td>0–10</td>
</tr>
<tr>
<td>Medium risk (30%–70% risk of delayed recovery)</td>
<td>11–14</td>
</tr>
<tr>
<td>High risk (&gt;70% risk of delayed recovery)</td>
<td>15+</td>
</tr>
</tbody>
</table>


Figure 2 Calibration plot for the final model. PPCS, persistent post-concussive symptoms.
significant correlation between female sex and neck tenderness at the initial clinical assessment (mean 6 days after injury), and untreated concomitant cervical injuries have been associated with longer recovery times.\textsuperscript{20} \textsuperscript{38} We typically treat neck injuries early according to recent guidelines.\textsuperscript{10} \textsuperscript{11} so it is not surprising that neck examination components (tenderness, spasm, reduced ROM) were not associated with PPCS risk in our sample. Future research should investigate the benefits of early management in those with neck injuries after concussion.

Time-since-injury was challenging to incorporate and, of the methods assessed, the bGLM with a complementary log–log link function performed best. One-point per day till presentation at the clinic was assigned in our scoring system, so children presenting 14 days after injury with ongoing signs and symptoms of concussion automatically qualified for the high-risk category for PPCS and were therefore candidates for PPCS planning. These findings are consistent with research by Kontos \textit{et al}.\textsuperscript{40} showing that recovery is delayed in those presenting to clinic later rather than sooner after injury. The increase in risk for PPCS for children presenting later could also be due to right censorship. Right censorship occurs when a subject is removed from the study before being included.\textsuperscript{41} So children who appear to \textit{wait a week and see if} they recover before being assessed by a physician but do not recover are at a higher risk for delayed recovery simply because children who recovered before being assessed by our clinicians were not included in the analysis. Further research investigating this relationship should be performed.

From the BCPE, the main effects of OI, VOR and tandem gait were included in the final predictive model. These elements assess the vestibular and autonomic,\textsuperscript{42} vestibular and ocular\textsuperscript{43} and dynamic vestibular\textsuperscript{44} subsystems, respectively. Impairments in each of these subsystems have been associated with longer recovery in concussed children,\textsuperscript{43} \textsuperscript{45} so we can be confident in including these in the BCPE RDR-score. While the majority of our subjects recovered from their vestibular deficits spontaneously, there is evidence that early vestibular rehabilitation may prevent persistent impairments;\textsuperscript{20} so, in certain cases of significant vestibular impairment, early directed intervention with a physical or occupational therapist should be considered as part of PPCS planning.

There are two interaction terms in our score, one positive and one negative. The negative interaction term between OI and VOR was included in both the Cox PH and bGLM models. Examining the coefficients for this term, when both OI and VOR were present, the HR in the Cox PH model and the combined coefficients in the bGLM model were comparable to when they were included separately. Hence, the combined effect of OI and VOR was not additive, and we can be confident in including a negative interaction term for their co-occurrence in the scoring system. The main effects of OI and VOR were highly significant (p<0.0005), while the coefficients for the OI and VOR interaction term were significant at a 0.05 level. We can therefore be confident that these predictors are valid; however, we are less certain about the main effect of tandem gait and the positive interaction term multiple/high-velocity injury and tandem gait. While these have been included in the model by stepwise selection using AIC, their individual coefficients had lower significance (p=0.259 and 0.103, respectively). A larger sample size would allow this issue to be resolved with more confidence.

Limitations

Our a priori sample size estimate did not account for unequal outcomes. With an average PPCS incidence of 30%, we should have collected a sample of 520 participants in total. Additionally, prior research has shown that generic rule-of-thumb sample size recommendations for multilevel logistic regression may not be reliable when distribution of predictors is not equal.\textsuperscript{46} A minimum recommended sample size for a future validation study based on 20 predictors-per-variable,\textsuperscript{47} 6 predictor variables from the BCPE RDR-score and 30 interaction pairs with a PPCS incidence of 36% would equal 2000 participants in total. Larger, multicentre studies should be performed to externally validate the integers of the scoring system.

Furthermore, our scoring system was developed using patients in a sports medicine clinic setting that incorporates early treatment for cervical injury; hence, this scoring system may not be predictive in children who are managed using older standards of care (ie, “cocooning”).\textsuperscript{48} Nevertheless, our clinic is not a “one stop shop” for concussion management and functions much as a primary care setting would in terms of referring patients to community providers for specialised treatment (eg, vision, vestibular, psychological and cognitive interventions). The majority of concussions in our sample were (1) from sport-related injuries (84%), (2) single/low-severity injuries (90%) and (3) in older children 12–18 years of age. We had few cases of assaults and MVAs (which are associated with much longer recovery times).\textsuperscript{7} These were classified as high-severity injuries, and almost all met criteria for high-risk for PPCS so we are confident about including them in the RDR-score; however, there were very few complex injuries in our sample (n=28), which increases the risk for a type 1 error.\textsuperscript{49} Future studies should validate the BCPE RDR-score in a more heterogeneous population. Our scoring system includes injury severity, which depends on an accurate patient history. This may not be reliable with children; hence, confirming details of the injury with observers is advised. Additionally, some SRC may not necessarily qualify as “low-velocity” and may be associated with additional trauma (eg, musculoskeletal injury) that may delay recovery, so each case should be classified according to all of the clinical information. Finally, this scoring system does not incorporate subjective concussion symptom checklists or mood-related diagnoses, which are known to be associated with delayed recovery.\textsuperscript{38} \textsuperscript{50} We did not include symptoms in our model because of concerns about under-reporting or over-reporting due to secondary gain motives.\textsuperscript{51} Thus, the RDR-score can be used in concert with symptom checklists, mental health and migraine history, and adjunct testing to enhance overall predictive performance.

CONCLUSION

The RDR-score is an easy-to-use, clinically relevant scoring system that, within 14 days of injury, can identify children at risk for developing PPCS. It is 85% accurate using 3 components of the history (≥3 previous concussions, days since injury and severity of injury) and 3 components of the BCPE (OI, VOR and tandem gait) to classify children at low (<30% risk, score=0–10), medium (30%–70% risk, score=11–14) or high (>70% risk, score=15+) risk for developing PPCS. Physicians in multiple settings can use this tool to implement early PPCS planning for high-risk patients, take a \textit{wait and see} approach for low-risk patients and decide on treatment recommendations based on patient-specific goals for children at medium-risk for developing PPCS. Early PPCS planning should be tailored to each patient based on underlying impairments, which are identified from a complete concussion evaluation consisting of
What are the findings?

- Using three components of the Buffalo Concussion Physical Examination and three components of the history, physicians can use the risk of delayed recovery (RDR)-score as a decision rule to identify children who would benefit most from early Persistent PostConcussion Symptom (PPCS) planning.
- The RDR-score classifies children within 14 days of injury into low-risk (<30%), medium-risk (30%–70%) or high-risk (>70%) for developing PPCS with 85% accuracy.
- This RDR-score was generated independent of symptom checklists, mood screeners or mental health history, so it can be used in concert with them without overlap in predictive capability.

How might it impact on clinical practice in the future?

- Interventions for concussion are usually delayed because the majority of impairments spontaneously recover within 1 month, however, 15%–30% of children take longer to recover and develop PPCS. Using the RDR-score decisions rule, clinicians can identify children who are at a high risk and begin planning for early intervention.
- Symptoms of PPCS are treated by identifying underlying symptom generators and providing directed interventions. Initial clinical evaluation of an acute concussion should include (1) a complete history including mental health history; (2) focused examination of the oculomotor, vestibular and cervical systems for common postconcussive impairments; (3) symptom checklists and (4) screens for changes in cognition and mood.
- Early planning for PPCS can include monitoring symptom exacerbators, communication with school personnel regarding accommodations, referral for specific oculomotor, vestibular and/or neurobehavioural interventions and sport-specific therapies.

history (including symptom checklists and behavioural screens), a concussion-focused physical examination and adjunct testing (eg, cognitive or exertional) where clinically indicated.

Twitter Barry Willer @bswlller

Contributors MNH, BW and JL designed the study; MNH, JL, SD, HNS, MSF and RKJ contributed to study design, acquisition of data and revising the clinical management protocols; MNH and AC performed all the statistical analysis; MNH, AC, JL and BW contributed significantly in preparing the manuscript; all authors contributed to manuscript writing and gave final approval before submission.

Funding Research reported in this publication was supported by the National Institute of Neurological Disorders and Stroke of the National Institutes of Health under award number 1R01NS094444-01 and by the National Centre for Advancing Translational Sciences of the National Institutes of Health under award number 1R01NS094444 and by the National Centre for Advancing Translational Sciences of the National Institutes of Health under award number UL1TR001412 to the University at Buffalo. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Disclaimer The authors are well known in the field because of the Buffalo Concussion Treadmill Test (BCTT) and Buffalo Concussion Protocol, and the Buffalo Concussion Physical Exam (BCPE) shares the same moniker. There are no monetary gains associated with this naming pattern.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. All data relevant to the study are included in the article or uploaded as supplementary information.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

ORCID iDs
Mohammad Nadir Haider http://orcid.org/0000-0001-7190-0004
John J Leddy http://orcid.org/0000-0002-0370-1289

REFERENCES

Original research


Supplement 1 for:

Derivation of the Buffalo Concussion Physical Examination Risk of Delayed Recovery (RDR) Score to identify Children at Risk for Persistent Post-Concussive Symptoms

Adam Cunningham, MA, MS; M Nadir Haider, MD, PhD

Other coauthors: John J Leddy, MD; Barry S Willer, PhD; Scott Darling, MD; Heidi Suffoletto, MD; Michael Freitas, MD and Rajiv Jain, MD

Table of Contents

Response Variables ................................................................. 2
Predictor Variables ............................................................... 4
Time-since-injury in Predictive Models ........................................ 12
Multicollinearity ................................................................. 13
Results of Predictive Models .................................................... 15
Development of the RDR-Score .................................................. 20
Cross-validation and Summary .................................................. 23

Implementation Note

All statistical analyses for this report were performed using the R Programming Language. Graphs were generated using the ggplot2 package, and data analysis performed using the flexsurv, muhaz, survival, and survminer packages. Since we are primarily interested in building a predictive model, we shall use the terms “predictor” and “response” throughout this report to refer to independent and dependent variables respectively. Given that there are 15 predictors, a traditional significance level of 0.05 is inappropriate, so we use a family-wide error rate of 0.05, and a Bonferroni correction to determine the significance level for each individual test. With 15 predictors, this yields a Bonferroni-corrected significance level of 0.05/15 = 0.0033. Cross validation is used to ensure that reported measures of model performance are not overstated. In particular, when variable selection techniques are used, cross validation is applied to the entire sequence of modeling steps including the selection of predictors.
1. Response Variables

There are two response variables for each participant:
1. Recovery Time (continuous)
2. Normal versus Delayed Recovery (binary)

Recovery Time

Recovery times are discrete variables, since values are discretized to the nearest day. For the purposes of this report however, Time to Recovery are considered as “suitably continuous”, and methods for analyzing continuous variables are used. Recovery times in the sample span a wide range, from a minimum of 6 days to a maximum of 268 days, so any error introduced in discretizing to the nearest day can be considered small relative to the magnitude of the variable.

As per clinical management protocols, examinations were conducted weekly for the first four weeks (the expected window of recovery), after which they were conducted every two weeks while the patient was referred for focused therapy. For patients with long recovery times, examinations could take place every four weeks as needed. This introduces an element of uncertainty as to exactly when “recovery” took place, with this uncertainty increasing for longer recovery times. Exact recovery times are therefore interval censored with recovery usually taking place sometime before being cleared for return-to-play.

However, the intention is for one of the models in this report to be used to develop a scoring system to predict the risk of persistent symptoms. If interval censoring is ignored, then the predicted recovery times from the model will be greater than the actual recovery times, having been biased high by the specific way in which the study was run with weekly examinations. The distribution of recovery times is shown in the figure below, binned into intervals of 5 days. The blue line shows a normal distribution fitted to the recovery times.

Figure 1.1: A histogram of recovery times shows a right-skewed distribution with a clear departure from normality

It can be seen in Figure 1.1 that recovery times are right-skewed, as is often found with time-to-event data. The mean recovery time is approximately 35 days, while the median is much less at 22 days, and there are also a small number of patients (17/270) with very long recorded recovery times of 100 days or more. It can also be seen that recovery times are far from normally distributed. This is confirmed by the Shapiro-Wilk test for normality ($p = 2.35 \times 10^{-23}$) that is statistically significant at a 0.05 level of significance.

In cases such as this where variables depart significantly from normality and equal variance, variable transformations can be explored to correct for these departures. A frequently used transformation for time-to-event data is to fit linear models to the logarithm of the response rather than directly to the response. This technique is used for the accelerated failure time models and is often found to correct for variance that increases with the magnitude of a variable. The blue line again shows a normal distribution fitted to the data.
We note in Figure 1.2 that visually the logarithms of the recovery times are much closer to normally distributed than are the recovery times. Although the departure from normality is still statistically significant at a 0.05 level of significance, with a p-value for the Shapiro-Wilk test for normality of $5.02 \times 10^{-7}$, this is however much less than that seen for the recovery times. For this reason, tests and models involving recovery times in the following chapters will use the logarithm of the recovery times rather than the recovery times directly.

**Normal versus Delayed Recovery**

For the purposes of this study, with a sample consisting entirely of adolescents, a classification of persistent symptoms or Persistent Post-Concussion Symptoms (PPCS) means that the time-to-recovery was greater than or equal to 30 days. Persistent symptoms are therefore a binary categorical variable. For descriptive purposes, “persistent symptoms” is described in this report as having the two levels “Yes” (for recovery times greater than or equal than 30 days), or “No” (for recovery times of 29 days or less). Of the 270 patients in this study, 98 (36.3%) displayed persistent symptoms. We can therefore set a lower bound on the accuracy of any predictive model of 64%, simply by predicting that no patients have persistent symptoms regardless of any information we have about the patient.
2. PREDICTOR VARIABLES

The sample data contains 15 predictor variables, which fall into three broad categories:

Demographic: Sex (binary), age (continuous), and previous concussions (discrete).

Injury: Time (in days) since injury when the first examination was conducted (continuous, discretized), and whether or not the injury was Single/Low-velocity Injury (binary).

Physical Examination: Results of the ten clinical tests from the Buffalo Concussion Physical Exam (binary), classified in each case by whether or not the test result is normal or abnormal.

We note that:
- Time-since-injury is not a covariate in the same sense as the other predictors, but is the time at which the physical examination predictors were measured. This will be discussed further in “Time since injury in Predictive Models”.
- The physical examination predictors are, strictly speaking, time-dependent covariates that were measured once at time-since-injury. This will also be discussed further.

1. Sex

Patient sex is represented as a binary variable, Male or Female. Descriptive statistics are shown below. We note that the majority of patients are male (62%), with 38% being female. Female patients have a slightly higher proportion of persistent symptoms (42%) than males (33%). However, the difference in proportions between male and female patients is not statistically significant, with a Pearson chi-squared p-value of 0.183. There is a small but statistically significant difference in mean log recovery times, with female patients having longer recovery times on average. The p-value for the two-sample t-test is 0.000273, which is statistically significant at a Bonferroni-corrected significance level of 0.00333. This is consistent with the literature on the effect of sex on recovery time.

Figure 2.1: Sex has a statistically insignificant association with persistent symptoms but is significantly associated with log recovery times at a Bonferroni-corrected level.

2. Age

It can be seen that the mean patient age is 14.9 years, with a median of 15.1 years. Patient ages span a range from 8.36 to 18.7 years, with most patients being of high school age. A linear regression model does not show age to be a statistically significant linear predictor of log recovery times. The p-value for a test of linear independence is 0.258, which is greater than a significance level of 0.05. A logistic regression model does not show age to be a statistically significant predictor of persistent symptoms either. The associated p-value is 0.179, which is also greater than a significance level of 0.05. Age is therefore unlikely to be included as a predictor in a scoring system for persistent symptoms.
3. Previous Concussions

The number of previous concussions is a discrete variable, with a minimum value in this data set of zero and a maximum value of seven. We note that the majority of patients (54%) have not had a previous concussion, and only a small proportion of patients (less than 6%) had three or more previous concussions. An increasing number of previous concussions is clearly associated both with a higher proportion of persistent symptoms and with longer recovery times. This is confirmed by the logistic regression test (p-value of 0.00907) and linear regression test (p-value of 0.024) for no association, both of which are statistically significant at a 0.05 level. The proportion of patients with persistent symptoms is similar for patients with 0, 1, or 2 previous concussions, only rising from 32% to 38%. Hence, 3 or more previous concussion was added as a predictor in the model.

For this new variable, the proportion of patients with persistent symptoms is statistically significantly different between patients with and without several previous concussions at a 0.05 level of significance. 35% of patients with 2 or less previous concussions displayed persistent symptoms, compared to 67% of patients with 3 or more previous concussions.
4. Single-impact/Low-velocity Injury or Multiple impact/High-velocity Injury

Single/Low-velocity or Multiple/High-velocity injury is a binary categorical variable and is obtained from the history provided to the physician who categorizes this injury. We note that most of the injuries were Single/Low-velocity Injury (n = 242), which is approximately 90% of the total. The Pearson chi-squared p-value is 0.00232, which is statistically significant at a Bonferroni-corrected level of significance of 0.00333. Recovery times are clearly longer on average for injuries that are Multiple/High-velocity Injury. This is confirmed by the two-sample t-test, which has a p-value of 1.93*10^{-7}. This is also statistically significant at a Bonferroni-corrected level of significance of 0.00333.

Figure 2.4: Single/Low-velocity Injury is significantly associated with both persistent symptoms and log recovery times at a Bonferroni-corrected level

5. Time-Since-Injury

Figure 2.5: Time since injury is significantly associated with both persistent symptoms and log recovery times at a Bonferroni-corrected level

Time-since-injury is measured in days from the initial injury until the first physical examination. We note that there is a clear positive correlation between time-since-injury and log recovery times. This is confirmed by the test for linear independence, which has a p-value of 2.3*10^{-15}. Time-since-injury is a
statistically significant predictor of persistent symptoms, with a logistic regression test of independence p-value of $3.49 \times 10^{-10}$.

6. Orthostatic Intolerance (OI)

The proportion of patients with persistent symptoms is statistically significantly different between patients with and without OI. 27% of patients without OI displayed persistent symptoms, compared to 48% of patients with it. The Pearson chi-squared test for independence has a $p$-value of 0.000607, which is less than a Bonferroni-corrected significance level of 0.00333. Mean log recovery times are also statistically significantly different between patients with and without OI. The two-sample t-test test of equal means has a $p$-value of 0.00176, which is also less than a Bonferroni-corrected significance level of 0.00333.

7. Neck Spasm

A small proportion of patients (7%) have abnormal neck spasm. The proportion of patients who displayed persistent symptoms is almost the same in patients with normal neck spasms (36%) as in patients with abnormal neck spasms (42%). The difference is not statistically significant, with a Pearson chi-squared $p$-value of 0.765. Mean log recovery times for patients with normal or abnormal neck spasms are also not statistically significantly different. The $p$-value of the two-sample t-test for equal means was 0.707.

8. Neck Tenderness

Approximately equal proportions of patients had normal (51%) and abnormal (49%) neck tenderness. The proportion of patients who displayed persistent symptoms is almost the same in patients with normal neck tenderness (35%) compared to abnormal neck tenderness (38%). The difference is not statistically significant, with a Pearson chi-squared $p$-value of 0.687. Mean log recovery times for patients with or without abnormal
neck tenderness are not statistically significantly different. The \( p \)-value of the two-sample t-test for equal means was 0.545.

9. Neck Range of Motion

A small proportion of patients (12%) have abnormal neck range of motion. The proportion of patients who displayed persistent symptoms is similar in patients with normal (38%) or abnormal (27%) neck range of motion. The difference is not statistically significant, with a Pearson chi-squared \( p \)-value of 0.338. Mean log recovery times are not statistically significantly different between patients with normal versus abnormal neck range of motion. The \( p \)-value of the two-sample t-test for equal means was 0.756.

Figure 2.8: Neck range of motion is not significantly associated with either persistent symptoms or log recovery times

10. Smooth Pursuits

The proportion of patients with normal smooth pursuits (41%) and abnormal smooth pursuits (59%) are fairly similar. A greater proportion of patients with abnormal smooth pursuits display persistent symptoms (44%) than do patients with normal smooth pursuits (26%). This difference is statistically significant at a Bonferroni-corrected significance level of 0.00333, with the Pearson chi-squared test for equal proportions having a \( p \)-value of 0.00325. Mean log recovery times are also statistically significantly different between patients with normal and abnormal smooth pursuits. The \( p \)-value of the two-sample t-test for equal means was 0.00222, which is also statistically significant at a Bonferroni-corrected significance level of 0.00333.

Figure 2.9: Smooth pursuits is significantly associated with both persistent symptoms and log recovery times at a Bonferroni-corrected level
11. Horizontal Repetitive Saccades

The proportion of patients with normal horizontal saccades (41%) and abnormal horizontal saccades (59%) was fairly similar. The proportion of patients with abnormal horizontal saccades who display persistent symptoms (43%) is greater than those with normal horizontal saccades (27%). This difference is not statistically significant at a Bonferroni-corrected significance level of 0.00333. At a 0.05 level of significance, mean log recovery times are statistically significantly different between patients with normal and abnormal horizontal saccades.

12. Vertical Repetitive Saccades

The proportion of patients with normal vertical saccades (46%) and abnormal vertical saccades (55%) was very similar. The proportion of patients with abnormal vertical saccades who display persistent symptoms (44%) is greater than those with normal vertical saccades (28%). This difference is statistically significant at a 0.05 level of significance, with a Pearson chi-squared test p-value of 0.00994, although this was not statistically significant at a Bonferroni-corrected significance level of 0.00333.

Figure 2.10: Horizontal and vertical repetitive saccades is significantly associated with both persistent symptoms and log recovery times at a 0.05 level

13. Vestibulo-Ocular Reflex (VOR)

The proportion of patients with normal (54%) and abnormal VOR (46%) was very similar. A greater proportion of patients with abnormal VOR display persistent symptoms (48%) than do patients with normal VOR (27%). This difference is statistically significant at a Bonferroni-corrected significance level of 0.00333, with a Pearson chi-squared test for equal proportions p-value of 0.00043. The p-value of the two-sample t-test for equal means was 0.00426, although this was not statistically significant at a Bonferroni-corrected significance level of 0.00333.

Figure 2.11: Vestibulo-ocular reflex is significantly associated with both persistent symptoms and log recovery times at a Bonferroni-corrected level
14. Near Point of Convergence (NPC)

A relatively small proportion of patients (20%) have abnormal NPC. The proportion of patients who displayed persistent symptoms is almost the same in patients with normal (36%) or abnormal (39%) NPC. The difference is not statistically significant.

Figure 2.12: Near point of convergence is not significantly associated with either persistent symptoms or log recovery times

15. Tandem Gait

A fairly similar proportion of patients have normal tandem gait (58%) as have abnormal tandem gait (42%). At a 0.05 level of significance, the proportion of patients with normal tandem gait who displayed persistent symptoms (31%) is not statistically significantly different to those with abnormal tandem gait (43%). The Pearson chi-squared test for equal proportions had a value of 0.068. Mean log recovery times for patients with normal or abnormal near point of convergence are not statistically significantly different. The p-value of the two-sample t-test for equal means was 0.065.

Figure 2.13: Tandem gait is not significantly associated with either persistent symptoms or log recovery times
Summary

There are very few missing values in the sample data, with only 7 values missing for OI, while the rest of the data is complete. Observations containing missing values are included when generating descriptive statistics for each predictor, but are removed when constructing predictive models. This yields a training set of 263 observations. The following graph summarizes the test for independence results for each predictor. P-values are plotted on a log scale, so the longer bars correspond to smaller p-values. At a 0.05 level of significance, age, convergence, neck range of motion, neck spasm, neck tenderness, and tandem gait have no statistically significant association with either log recovery times or with persistent symptoms. They are therefore unlikely to be good predictors of persistent symptoms.

Figure 2.14: Tests of independence between predictor and response variables
3. TIME-SINCE-INJURY IN PREDICTIVE MODELS

Time-since-injury was the predictor most significantly associated with both persistent symptoms (p = 3.49*10^{-10}) and log recovery times (p = 2.3*10^{-15}). Representing this information correctly in predictive models is therefore critical to model fit and performance. The simple linear regression model below predicts recovery time from time-since-injury, and shows the strong association that exists between the two variables.

Figure 3.1. Linear regression model predicting recovery time from time-since-injury

The red line shows the fitted model, the gray band a 95% confidence interval for the fitted regression line, and the dashed black line the lower limit for recovery time (i.e. time-since-injury). The model has the approximate form: Recovery time = 14 + 3.5*time-since-injury (in days)

Time-since-injury in Cox PH Models

Time-since-injury represents a delayed entry time, and would seem to be represented most naturally in time-to-event models as a left-truncation point. A key assumption however when using left-truncation in Cox PH models is that the event times and the delayed entry times are independent, given the covariates. Since recovery times are strongly dependent on time-since-injury, an alternative approach is to include time-since-injury directly as a predictor in Cox PH models. To explore whether hazard ratios are constant across time for patients with different time-since-injury, we plot estimated hazard rates over time for patients in three groups: time-since-injury between 0 and 5 days (148 patients), between 6 and 10 days (83 patients), and greater than 10 days (32 patients). Estimated hazard ratios for patients with different time-since-injury are were not constant over time. Hence, a stratified Cox PH Model will be used going forward.

Time-Since-Injury in Accelerated Failure Time Models

As with Cox PH models, a theoretically well motivated way to represent time-since-injury is as a delayed entry time. Hence, a question of interest is whether any AFT models could adequately model the relationship seen in the sample data between recovery times and time-since-injury through left-truncation alone. The Weibull and log-logistic distributions both have closed-form survival functions, so the relationship between time-since-injury and median recovery times can be calculated analytically for these models.

Time-Since-Injury in Binomial Generalized Linear Models

Since there is a one-to-one correspondence between AFT models and bGLMs, different choices of error distribution for AFT models yield different link functions for their associated binomial GLMs, while the linear predictor for a bGLM is a linear transformation of that for the associated AFT model. Representing time-since-injury as a predictor in AFT models therefore naturally leads to also representing time-since-injury as a predictor in binomial GLMs.
4. MULTICOLLINEARITY

We note that multicollinearity is a function of the design matrix alone (i.e., just the predictor data), and not of the response. Before we develop predictive models, it is therefore worth examining this aspect of the data to see if any problems may occur. Given the nature of the predictors, which include multiple clinical tests of the same neurological systems, it is reasonable to assume that correlations between predictors will exist. Two primary ways to reduce the effects of multicollinearity are:

1. Remove highly correlated predictor variables from the model.
2. Use partial least squares regression or principal component analysis to reduce the number of predictors to a smaller, uncorrelated set.

Independence Tests Between Predictor Variables

Figure 4.1: Tests of independence between predictor variables

Every predictor variable is plotted against every other predictor, with colored bars indicating proportions of abnormal test results, or male patients in the case of sex. Labels on the diagonal refer to the x-axis in the same column and the y-axis in the same row. P-values for tests of independence are color-coded on each graph. Red denotes significance at a Bonferroni-corrected level, corrected for 120 pairwise tests.
We observe the following:

1. Age and several previous concussions are not significantly associated with any other predictors at a 0.05 level of significance.
2. Convergence is not significantly associated with most other predictors at a 0.05 level of significance, and with no other predictors at a Bonferroni-corrected level of significance.
3. Single/Low-velocity injury and time-since-injury have little significant association with most other predictors. This is particularly relevant for time-since-injury, as it means that the symptoms for patients entering the study a longer time after their injury are neither better nor worse than those for patients whose injury was more recent. This is explored in more detail later.
4. Sex is not significantly associated with most other predictors, with the main associations being with OI, single/low-velocity injury, and time-since-injury.
5. Horizontal saccades, vertical saccades, smooth pursuits and VOR form a cluster of related predictors, with each predictor significantly associated with the others at a Bonferroni-corrected level of significance.
6. Tandem gait is less strongly related to this cluster, being significantly associated with vertical saccades and VOR at a Bonferroni-corrected level of significance, and with horizontal saccades and smooth pursuits at a 0.05 level of significance.
7. OI is further out of this cluster, being significantly associated with VOR at a Bonferroni-corrected level of significance, and to horizontal saccades, vertical saccades, and tandem gait at a 0.05 level of significance, but not with smooth pursuits.
8. Neck tenderness is significantly associated with both neck range of motion and neck spasms at a Bonferroni-corrected level of significance, although neck range of motion and neck spasms are not significantly associated with each other.

**Time-Dependent Covariates**

We are aware that the predictors from the initial physical examination are in fact *time-dependent covariates*. Whereas the values of the predictors age, sex, several previous concussions and type of injury are fixed at the time of the concussive injury, the results of the physical examination tests are dependent on the recovery process and can be expected to vary depending on how long after the injury the initial examination is performed. Time-dependent covariates are known to cause problems with regression models in several ways. Firstly, it is far from obvious how to model the functional form of a time-dependent covariate, and doing so introduces new complexity to models with an associated risk of overfitting. Secondly, with time-dependent covariates, the ability to predict in Cox regression models is usually lost. The existence of a measurement at a particular time implies that a subject is still in the study, and has not yet recovered. This may not be a particular issue however, given that tests are only performed once at entry to the study. Since left-truncation occurs at the same time as tests are conducted, the inclusion of patients in the study at this time is already explicitly modeled using left-truncation. This is a difficult issue to deal with rigorously in time-to-event models. Correlation between physical exam components and time-since-injury was analyzed and no predictor was significantly time dependent after correction. In the absence of any clear techniques to incorporate this information apart from the use of left-truncation, we do not attempt to explicitly model time-dependent covariates in this analysis.
5. RESULTS OF PREDICTIVE MODELS

Results of each model fit are presented in this section. The best performing model out of all classes of model will then be used to develop a scoring system for persistent symptoms. Leave-one-out cross validation is used on all models to estimate performance on new data. Unless otherwise indicated, references to model performance refer to cross validated results rather than results on training data.

Model Coefficients for Models Based on All Predictors

Figure 5.1: Model coefficients for models based on all predictors
It can be seen that for models which include time-since-injury as a predictor (the AFT and bGLM), this is the most significant predictor of persistent symptoms at a Bonferroni-corrected level of significance. Orthostatic intolerance and single/low-velocity injury are also statistically significant predictors in all three models at either a Bonferroni or a 0.05 level of significance. The bGLM with a corresponding log-log link function has the greatest number of statistically significant predictors, while the Cox PH model has the least. A reason may be that stratifying by time-since-injury in the Cox PH model reduces power by reducing the number of observations available to estimate the hazard rate for each stratum.

**Model Coefficients for Stepwise-Selected Models Based on All Predictors**

*Figure 5.2: Coefficients for stepwise-selected models based on all predictors*

It can be seen that the number of predictors in these models is less than half that of the full models. These models would therefore result in a simpler scoring system. The most significant predictors in the full models of the previous section have all been retained in the stepwise selected models (time-since-injury, OI, simple injury), but the order of the less significant predictors has changed slightly as predictors have been dropped from the models. Although not shown here, the number of predictors in the AFT models with higher AIC (log-logistic, lognormal and Weibull) was greater (7 predictors in each model).

**Interaction Terms**

There are several reasons for considering pairwise interaction terms between the predictors. The first is that statistically significant interactions between predictors may well exist. There are however pitfalls involved with using interaction terms, particularly the risk of overfitting the data. With 15 predictors there are 105 pairwise interaction terms, and only 270 observations in the sample data. Including all these terms in a model would severely violate the suggestion that the number of observations used to fit a regression model should exceed the number of predictors (or candidate predictors if using variable selection) by a factor of at least ten.
Figure 5.3: Interaction terms between predictor variables in two-variables models

Significance is shown using both a 0.05 level of significance and a Bonferroni-corrected level of significance of 0.05/105 = 0.000476 since there are 105 separate interaction terms being tested.

It can be seen that the only interaction terms that are statistically significant at a 0.05 level of significance are those involving VOR/OI, and sex/neck tenderness. No interaction terms are significant at a Bonferroni-corrected level of significance. This is an interesting result given that OI and VOR were both found to be individually significant predictors of persistent symptoms. The subsequent models will include all predictors plus all pairwise interaction terms. This makes 105 variables in total for the Cox PH model (14 main effects plus 91 interaction terms), and 120 variables for the AFT and binomial GLMs which include time-since-injury as a predictor (15 main effects plus 105 interaction terms). Stepwise selection using the AIC is then employed to see if any interaction terms are considered sufficiently predictive to be included.
Model Coefficients for Stepwise-Selected Models Based on All Interactions

Figure 5.4: Coefficients for stepwise-selected models based on all pairwise predictor interactions

It can be seen that very few interaction terms are included in the stepwise selected models. This is expected given that only two interactions were individually significant at a 0.05 level when included in two-variable models. Only two interaction terms have been included in the Cox PH model and binomial GLM, and none in the inverse-Weibull model. Although not shown, these models have less terms than the other AFT models (7 in the log-logistic AFT, 9 in the lognormal AFT, and 13 in the Weibull AFT), and in the other binomial GLMs (15 in the models with logit and probit links, and 12 in that with a cloglog link). Hence, the models shown are the simplest in their class.

Summary

The table below summarizes the results of modeling. Pairs of numbers in a cell refer to performance on training and test data, with the upper number showing performance on the training data, and the lower number the performance using cross-validation. For AFT and binomial GLM models, results are only given for the model with the lowest AIC. This was the inverse-Weibull in every case, and the binomial GLM with a corresponding log-log link in every case except the stepwise-selected model include all pairwise predictor interactions.
### Table 5.1: Summary of model performance

<table>
<thead>
<tr>
<th>Model (best AFT)</th>
<th>Binomial GLM (Clog-log)</th>
<th>AFT Model (Inverse-Weibull)</th>
<th>Cox PH Model (Stratified)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accuracy (%)</td>
<td>Sensitivity (%)</td>
<td>Specificity (%)</td>
</tr>
<tr>
<td>Models with Main Effects Only</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Null model</td>
<td>64</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>Demographic</td>
<td>73</td>
<td>47</td>
<td>87</td>
</tr>
<tr>
<td>Exam</td>
<td>67</td>
<td>35</td>
<td>86</td>
</tr>
<tr>
<td>All predictors</td>
<td>76</td>
<td>54</td>
<td>89</td>
</tr>
<tr>
<td>Models Including Interaction Terms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All interactions</td>
<td>92</td>
<td>87</td>
<td>96</td>
</tr>
<tr>
<td>Stepwise Selected Models</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All predictors stepwise</td>
<td>77</td>
<td>57</td>
<td>89</td>
</tr>
<tr>
<td>All interactions stepwise</td>
<td>78</td>
<td>56</td>
<td>91</td>
</tr>
</tbody>
</table>

### Binomial GLM Assumptions

Before we use this model to devise a scoring system, we perform the Hosmer-Lemeshow test of goodness of fit commonly used in regression models.

**Figure 5.5: Binomial GLM test of goodness of fit**

The p-value for the Hosmer-Lemeshow tests is greater than 0.05 for all binomial GLMs, hence we can be confident that our model is not violating assumptions of logistic regression.
6. DEVELOPMENT OF THE RDR-SCORE

The following table summarizes the details of the “best” model.

**Table 6.1: Details of “best” model**

| Variable                                      | Coeff.    | Std. Err. | z       | P > |z|    | 95% CI               |
|-----------------------------------------------|-----------|-----------|---------|-----|-----|----------------------|
| Time-since-injury                             | 0.2585    | 0.0363    | 7.125   | 1.04 × 10ⁱ² | 0.1874 | 0.3296               |
| Orthostatic intolerance                       | 1.3051    | 0.3716    | 3.512   | 0.0004 | 0.5768 | 2.0333               |
| Vestibulo-ocular reflex                       | 1.2874    | 0.3714    | 3.466   | 0.0005 | 0.5595 | 2.0154               |
| > 2 previous concussions                      | 1.0754    | 0.4545    | 2.366   | 0.0180 | 0.1846 | 1.9663               |
| Tandem gait                                   | 0.2931    | 0.2596    | 1.130   | 0.2586 | -0.2155 | 0.8018              |
| High/multiple impact injury                   | 0.5003    | 0.4724    | 1.059   | 0.2896 | -0.4256 | 1.4262              |
| OI*VOR                                        | -1.1326   | 0.4969    | -2.279  | 0.0226 | -2.1064 | -0.1687              |
| High impact*tandem gait                       | 1.2672    | 0.7760    | 1.633   | 0.1025 | -0.2538 | 2.7882              |
| Intercept                                     | -3.6921   | 0.4106    | -8.991  | < 2 × 10⁻¹⁶ | -4.4970 | -2.8872           

The following transformations are made to form a simpler and more intuitive system:

1. Integer coefficients: A scoring system based on integer values would be simplest to use and understand. The main issue here is transforming the model coefficients to a set of integers while minimally changing the model predictions.
2. Positive coefficients: The coefficient for having a single/low-velocity injury is negative in the binomial GLM. A scoring system should reverse this so that the reference level for each predictor is the lowest risk level.
3. Zero minimum: The intercept term means that the linear predictor does not equal zero when all predictors have a value of zero. A minimum value of zero for the scoring system would be easier to understand, and correspond to the lowest possible risk.

To convert all coefficients for main effects to positive values, a new model was generated with single/low-velocity injury as the reference level. The new coefficients are shown below, with “multiple/high-velocity” being the opposite of “single/low-velocity”.

**Figure 6.1: Predictors chosen for the stepwise-selected corresponding log-log binomial GLM based on all predictor interactions with minimum risk reference levels**

An integer-based scoring system can be created by scaling all estimated coefficients by a multiplier \( m \). The table below shows the scores assigned to each predictor for each value of \( m \). The last two rows show the root mean squared error per predictor and the maximum score for each system.
Table 6.2: Integer scores for each predictor based on different values of the multiplier m

<table>
<thead>
<tr>
<th>Predictor multiplier m</th>
<th>2.487</th>
<th>3.832</th>
<th>6.277</th>
<th>10.233</th>
<th>14.055</th>
</tr>
</thead>
<tbody>
<tr>
<td>OI</td>
<td>3</td>
<td>5</td>
<td>8</td>
<td>13</td>
<td>18</td>
</tr>
<tr>
<td>VOR</td>
<td>3</td>
<td>5</td>
<td>8</td>
<td>13</td>
<td>18</td>
</tr>
<tr>
<td>Over 2 previous concussions</td>
<td>3</td>
<td>4</td>
<td>7</td>
<td>11</td>
<td>15</td>
</tr>
<tr>
<td>Tandem gait</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Multiple/high velocity injury</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>VOR*OI</td>
<td>-3</td>
<td>-4</td>
<td>-7</td>
<td>-12</td>
<td>-16</td>
</tr>
<tr>
<td>High velocity*tandem gait</td>
<td>3</td>
<td>5</td>
<td>8</td>
<td>13</td>
<td>18</td>
</tr>
<tr>
<td>RMS error</td>
<td>0.056</td>
<td>0.021</td>
<td>0.013</td>
<td>0.008</td>
<td>0.005</td>
</tr>
<tr>
<td>Maximum score</td>
<td>11</td>
<td>18</td>
<td>29</td>
<td>46</td>
<td>64</td>
</tr>
</tbody>
</table>

The decrease (on average) in the root mean squared error per predictor as $m$ increases shows that, if we allow $m$ to be arbitrarily large, the transformed values can be made as close to the original as desired. The disadvantage of increasing $m$ is that the scores also become larger, leading to a more complicated scoring system. Ideally, the value of $m$ should be chosen so that (i) the estimated probabilities calculated using the scoring system are close to those using the binomial GLM, and (ii) the integers used in the scoring system are low. To check (i), we plot below the predicted estimated using the binomial GLM against scores calculated on the sample data. The actual outcomes (persistent/non-persistent symptoms) for each patient are color coded.

Figure 6.2: Estimated probabilities of persistent symptoms plotted against scores for scoring systems based on different values of the multiplier m.

It can be seen that when $m = 3.832$, the score coefficient of 0.9905 for time-since-injury is fortuitously close to the integer 1. Using this value of $m$ would therefore mean that time-since-injury would not need to be multiplied by a non-integer value, but could simply be included in the score without modification. Assigning a score to a patient is simply done by summing the scores for each predictor. The interaction terms in this scoring system can be thought of as adjustments for additional or less risk.
Table 6.3: Final integer-based scoring system for estimating the risk of Delayed Recovery

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time-since-injury</td>
<td>1 point per day</td>
</tr>
<tr>
<td>High-velocity/multiple impact</td>
<td>2 points</td>
</tr>
<tr>
<td>More than 2 previous concussions</td>
<td>4 points</td>
</tr>
<tr>
<td>OI</td>
<td>5 points</td>
</tr>
<tr>
<td>VOR</td>
<td>5 points</td>
</tr>
<tr>
<td>Tandem gait</td>
<td>1 point</td>
</tr>
<tr>
<td>OI and VOR</td>
<td>-4 points</td>
</tr>
<tr>
<td>High-velocity/multiple impact and tandem gait</td>
<td>5 points</td>
</tr>
</tbody>
</table>

The final part of the scoring system is deciding how to map scores to risk categories. Overall accuracy is dependent on which scores are chosen as cutoffs for the risk categories. The binomial GLM had 78% training accuracy, with 51% sensitivity and 90% specificity using a cutoff probability of 0.5. Choosing a cutoff of 0.3 for low risk decreases overall accuracy and specificity, but increases sensitivity considerably (from 51% to 72%). From the graph, it can be seen that most patients with a score of 10 or less (corresponding to a probability of 0.30) did not develop persistent symptoms, while most patients with a score of 15 or greater did (corresponding to a probability of 0.70). Patients with scores from 11 to 14 appear to be a mixture of those who did and did not develop persistent symptoms. This suggests that we can classify patients into low-, medium-, or high-risk categories using scores of 10 and 15 to mark the boundaries of low/medium and medium/high risk respectively.

Figure 6.3: Classification of patients using the scoring system based on all training data

<table>
<thead>
<tr>
<th>Score</th>
<th>Low risk</th>
<th>Medium risk</th>
<th>High risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>No persistent symptoms</td>
<td>122 (73%)</td>
<td>38 (23%)</td>
<td>6 (4%)</td>
</tr>
<tr>
<td>Persistent symptoms</td>
<td>27 (28%)</td>
<td>27 (28%)</td>
<td>43 (44%)</td>
</tr>
</tbody>
</table>

The proportions in each risk category were checked and found to be almost identical to those obtained using the estimated probabilities generated by the binomial GLM using cutoff probabilities of 0.3 and 0.7 (proportions were identical in the low category, and only differed by one or two patients in the medium and high-risk categories). This suggests that any discretization errors introduced by using the scoring system are minimal.
7. CROSS-VALIDATION AND SUMMARY

The first step in this process checks that the predictors chosen for the model trained on all the data are the same ones that would be chosen using leave-one-out cross validation. This was done by counting the number of stepwise-selected models that included each predictor (and their interaction pairs) using leave-one-out cross validation. We note that the OI*VOR interaction term was counted as either OI*VOR or VOR*OI, and was actually selected for every model. The predictors used to develop the scoring system are present in nearly every model. The exception is the multiple/high-velocity*tandem gait interaction term, which was not selected using the AIC in just 2 models (less than 1%). A small number of models (13% or less) included predictors not present in the scoring system, but these are all of low statistical significance. It would have been of more concern if the scoring system included predictors which only appeared in a few stepwise selected models, which did not occur. It seems therefore that stepwise selection of predictors for the model is reasonably robust, and that the predictors used in the scoring system have been validly included.

The next step is to calculate the scores that would be generated using leave-one-out cross validation for each model. Ideally the variation in the maximum likelihood estimates for the model coefficients will be small enough that the rounding process will lead to these matching the scores generated using all the sample data, but again this is a matter for experimental verification. The scoring system therefore seems relatively robust. The main reason for this is that the rounding process used to convert coefficients to integer scores is tolerant of coefficient changes that are small relative to the magnitude of the coefficient. To calculate the cross validated scores we generate a scoring system for every patient on a training set with the patient removed, then scored the patient according to that system. The scoring system is generated using stepwise-selected models based on all predictor interactions, so the models may differ in the predictors chosen. Scores for each predictor are found by dividing by the coefficient for time-since-injury and rounding, which fixes the score for time-since-injury at one for every model.

**Figure 7.1. Cross-validation of the BCPE RDR-Score using Leave-one-out method**

![Graph showing cross-validation results for persistent symptoms](image)

<table>
<thead>
<tr>
<th>Score</th>
<th>Low risk</th>
<th>Medium risk</th>
<th>High risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>No persistent symptoms</td>
<td>115 (69%)</td>
<td>42 (25%)</td>
<td>9 (5%)</td>
</tr>
<tr>
<td>Persistent symptoms</td>
<td>32 (33%)</td>
<td>28 (29%)</td>
<td>37 (38%)</td>
</tr>
</tbody>
</table>

For the model tested on new data using cross validation, 69% of patients without persistent symptoms fall into the low-risk category, and 67% of patients with persistent symptoms fall into either the medium- or high-risk categories. Performance on test data is therefore somewhat lower than on training data.

**Conclusions**

The goal of this study was to develop a scoring system to quantify the risk of persistent symptoms for adolescents presenting within 14 days of injury. The strategy adopted has been to fit predictive statistical
models to the patient data collected by the University at Buffalo Concussion Management Clinic, then use the parameters of the model to develop the scoring system.

Two factors have complicated the modeling process when compared to typical regression modeling: the presence of a delay between the concussive injury and the initial physical examination (referred to as time-since-injury), and the presence of time-dependent covariates in the results of the initial physical examination. The approach taken in this report has been to “let the data decide” how to incorporate this information into models. Analysis of the effect of these factors on the responses justified including them as predictors, which has allowed simpler and more understandable models to be developed.

The existence of a long tail in the distribution of recovery times has also been the subject of extensive analysis. Models based on the right-skewed extreme value (maximum) distribution were found to best accommodate this long tail. The inverse-Weibull AFT model and binomial GLM with a corresponding log-log link function, which are based on this distribution, were found to provide the best fit and simplest stepwise selected models for recovery times and persistent symptoms respectively.

A scoring system has been developed based on this binomial GLM that is simple, easy to use, and easy to understand. It is employs three questions about a patient’s injury and concussion history, and three clinical tests. These tests and demographic characteristics make sense clinically and statistically, and for the most part are the same predictors that would be selected by a simple univariate analysis of statistical significance.
Supplement 2 for:

**Derivation of the Buffalo Concussion Physical Examination Risk of Delayed Recovery (RDR) Score to identify Children at Risk for Persistent Post-Concussive Symptoms**

M Nadir Haider, MD, PhD; Barry S Willer, PhD; John J Leddy, MD

Other coauthors: Adam Cunningham, MA, MS; Scott Darling, MD; Heidi Suffoletto, MD; Michael Freitas, MD and Rajiv Jain, MD

Clinical Scenarios

**Scenario 1:** 15-year-old female hockey player presented with symptoms of headache, dizziness and difficulty concentrating for 3 days after falling and striking the back of her helmeted head during practice. She did not lose consciousness but was transiently confused immediately after the injury. Despite symptoms, she continued to play for another 20 minutes without getting another head injury. She has a history of one other hockey concussion a year ago that resolved within 2 weeks. On physical examination, there was no lightheadedness or dizziness on postural change (i.e., no orthostatic intolerance), she had cervical muscle tenderness on palpation but no loss of cervical range of motion, she experienced dizziness with horizontal VOR, and had 3 missteps on tandem gait with eyes closed. No impairments on visual tracking were observed. Remainder of clinical assessment was within normal limits.

<table>
<thead>
<tr>
<th>BCPE Component</th>
<th>Result</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Days since injury</td>
<td>3 days at 1-point per day</td>
<td>3</td>
</tr>
<tr>
<td>High velocity/Multiple impact</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>≥ 3 previous concussions</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>OI</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>VOR</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Tandem Gait</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>OI and VOR</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>High velocity/Multiple impact and Tandem Gait</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total Score</strong></td>
<td></td>
<td><strong>9</strong></td>
</tr>
</tbody>
</table>

Outcome: This patient qualifies as low risk for PPCS so early therapeutic intervention (other than general advice on relative rest and avoidance of risky activity) is not warranted; however, her cervical strain should be managed. There is minimal requirement to assist the child and parent to plan for delayed recovery.

**Scenario 2:** 12-year-old male presents with symptoms of headache, dizziness and difficulty concentrating for 10 days after falling from the upper portion of a bunk bed and hitting the side of his head on the floor. He did not lose consciousness but experienced headache and dizziness that got worse over the day. He has no history of prior concussion. On physical examination, he reported feeling lightheaded on postural change and was symptomatic during repetitive saccades and the VOR. He had a normal neck examination and a normal tandem gait. There was no suspicion of child abuse.
### BCPE Component Result Score

<table>
<thead>
<tr>
<th>BCPE Component</th>
<th>Result</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Days since injury</td>
<td>10 days at 1-point per day</td>
<td>10</td>
</tr>
<tr>
<td>High velocity/Multiple impact</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>≥ 3 previous concussions</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>OI</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>VOR</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Tandem Gait</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>OI and VOR</td>
<td>1</td>
<td>-4</td>
</tr>
<tr>
<td>High velocity/Multiple impact and Tandem Gait</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total Score</strong></td>
<td></td>
<td>16</td>
</tr>
</tbody>
</table>

Outcome: This patient qualifies for high risk for PPCS. He and his family would benefit from PPCS planning such as instruction on sleep hygiene and strategies for communication with school personnel regarding academic accommodations. If additional vestibular and oculomotor findings are present, consideration should be given to earlier institution of physical therapy.

### Cervical Management Strategies
Cervical injuries were managed by physicians at our practice using the following algorithm: Mild cervical injuries are managed conservatively by the physicians by prescribing over-the-counter analgesics, warm/cold compresses or self-regulated stretching exercises. Additionally, several children with sport-related injuries had access to their school or team’s athletic trainer/coach who helped supervise their neck stretches. However, if the cervical injury was moderate or higher and was causing dysfunction, then patients were also assessed by a physical therapist at the initial clinic encounter who provided specific cervical interventions in addition to their concussion management.

### Directions for Physical Examination Techniques

**Orthostatic Intolerance:** OI is measured using the 1-Minute Supine to Standing Orthostatic Hypotension (OH) test. The patient is asked to lay in a supine position for at least 2 minutes and a manual or automated blood pressure cuff is used to measure heart rate and blood pressure. The patient is then asked to stand up without support and with both feet firmly on the ground and a second measurement is taken after standing for 1 minute. The patient is asked if any increase in dizziness, lightheadedness or headache is experienced upon standing or by 1-minute. Patient is positive for OI if there is an increase in dizziness, lightheadedness or headache upon standing which goes away upon recumbency. Patient is positive for OH if there is a >20 mmHg drop in systolic or a >10 mmHg drop in diastolic blood pressure in the standing position when compared to supine values.

**Vestibular-Ocular Reflex:** To assess the VOR, place your thumb about a foot from the patient’s face and ask the patient to focus on the thumb. Have the patient rotate the head side to side as fast as is comfortable while keeping fixation on the thumb for at least 10 repetitions while you look at their eyes. The VOR is considered abnormal if there is slow movement or loss of fixation on the thumb (eyes beating back to the center) and/or exacerbation of headache, dizziness, nausea, or blurred or double vision on motion.

**Tandem Gait:** The patient is asked to walk in a straight line for 5 steps, heel to toe, with hands at the side, while looking straight ahead on a fixed point on the wall. The patient then walks backwards, toe to heel, along the same line while looking straight ahead. The patient then performs the tandem gait again with eyes closed. Tandem gait is considered abnormal if the patient is unable to walk the line, stumbles, steps out of line or has to open their eyes to regain balance during the eyes closed portion of the test. However, minimal imbalance with the eyes closed tandem gait without any abnormalities in the eyes open portion is considered normal.