Concussion Symptom Inventory (CSI):
An Empirically-Derived Scale for Monitoring Resolution of Symptoms Following
Sport-Related Concussion

Christopher Randolph, PhD
William B. Barr, PhD
Michael McCrea, PhD
Scott Millis, PhD
Kevin Guskewicz, PhD, ATC
Thomas A. Hammeke, PhD
James P. Kelly, MD

1. Department of Neurology, Loyola University Medical Center, Chicago (Maywood), IL
2. Departments of Neurology and Psychiatry, New York University School of Medicine, New York, NY
3. Neuroscience Center, Waukesha Memorial Hospital, Waukesha, WI
4. Department of Neurology, Medical College of Wisconsin, Milwaukee, WI
5. Department of Physical Medicine and Rehabilitation, Wayne State University School of Medicine, Detroit, MI
6. Departments of Exercise and Sport Science and Orthopedics, University of North Carolina at Chapel Hill, Chapel Hill, NC
7. Department of Neurosurgery, University of Colorado School of Medicine, Denver, CO

Correspondence to: Christopher Randolph, PhD, 1 East Erie, Suite 355, Chicago, IL 60611.  tel: (312)-863-3033; fax (312)-573-0900; email crandol@lumc.edu.

Acknowledgements:  This project was supported in part by funding from the NCAA, the National Operating Committee on Standards for Athletic Equipment (NOCSAE), Center for Disease Control’s National Center for Injury Prevention and Control (NPIPC), National Academy of Neuropsychology, Waukesha Memorial Hospital Foundation, National Federation of State High School Associations, NFL Charities, Green Bay Packer Foundation, Milwaukee Bucks, Herbert H. Kohl Charities, Waukesha Service Club, and the Medical College of Wisconsin General Clinical Research Center (M01-RR00058 from the National Institutes of Health). The authors would also like to acknowledge the invaluable assistance of Amy Mathews, MSW, and Stephen Marshall, PhD, in data management.
**Objective:** Self-report post-concussion symptom scales have been used for many years as a key method of monitoring recovery from sport-related concussion, to assist in medical management and return-to-play decision-making. To date, however, item selection and scaling metrics for these instruments have been based solely upon clinical judgment, and no one scale has been identified as the “gold standard”. The goal of this project was to use a statistical approach to explore item selection and scaling from a large dataset of existing scales, in order to empirically-derive the most efficient and appropriate scale possible for this application. **Setting:** Data were collected as part of three separate studies of sport-related concussion, involving 129 high schools and 29 colleges. **Participants:** Baseline data from detailed standardized symptom checklists including a total of 27 symptom variables were collected from a total of 16,350 athletes, including 13,879 male and 2,471 female participants in football, soccer, lacrosse, and ice hockey. Follow-up data were obtained from 641 athletes who subsequently incurred a concussion. **Main Outcome Measurements:** Symptom checklists were administered at baseline (preseason), immediately post-concussion, post-game, and at 1, 3, 5, and 7 days post-injury. **Results:** Effect size and Rasch analyses resulted in retention of only 12 of the 27 variables, and in a change from Likert to dichotomous scaling. Receiver-operating characteristic (ROC) analyses were used to confirm that the reduction to dichotomous scaling did not reduce sensitivity or specificity. The newly-derived Concussion Symptom Inventory (CSI) is presented. **Conclusions:** Using an empirical approach to eliminate items that proved to be insensitive to concussion and to reduce the score range from 7 to 2 resulted in a scale that is efficient and rapidly-administered, without sacrificing sensitivity or specificity. The CSI is recommended as a core measure for monitoring recovery from sport-related concussion.
Introduction

The medical management of sport-related concussion has suffered from a dearth of empirical data from prospective controlled outcome studies. This has led to a burgeoning number of conflicting injury classification systems and return-to-play guidelines. Although there are now over a dozen different proposed sets of guidelines, it has been recognized that few, if any, of these are evidence-based, and none has been universally accepted. The various guidelines are all in agreement, however, that a player should be symptom-free before returning to play.

Although the rationale for this recommendation also remains poorly substantiated to date, the primary concern is that players may be at an elevated risk of repeat concussion during the symptomatic post-concussive period. There is some evidence that such a period of vulnerability may exist, and that recovery following a second concussion may be somewhat more prolonged. A second concern is the risk of “second-impact syndrome,” or brain swelling thought to be secondary to cerebrovascular congestion. This can be a life-threatening condition, but it is extremely rare and the causative mechanism remains unclear (and may not require a “second” impact).

There is a general consensus, however, that until these risks are clarified, concussed players should be free of symptoms before return to competition. A number of methods have been explored to measure concussion-related symptoms or impairments, including brief “sideline” neurocognitive examinations, balance testing, and more extensive neuropsychological testing, designed to detect changes in cognitive functioning by comparing players to their own preseason baseline. The use of self-report subjective symptom checklists or scales has also been a consistent component of concussion management, and these have repeatedly been demonstrated to be sensitive to the effects of concussion.

Self-report symptoms are also the primary decision-making factor in the most commonly used guidelines for return to play. Concussed athletes typically show elevated scores on symptom concussion checklists for at least as long as impairment is detectable via more time-consuming and expensive methodologies (e.g., neuropsychological testing), despite concerns that players might underreport symptoms in order to be cleared to return to play. Finally, recent publications, including a consensus paper, have recommended that players should be asymptomatic before screening for impairment using any type of neuropsychological testing, further underscoring the central role of subjective symptom checklists in monitoring recovery from concussion.
A variety of subjective symptom scales have been used in the study of sport-related concussion, although these typically involve substantial overlap in item content. Items are typically chosen on the basis of clinical experience with concussion-related symptoms, and the overall sensitivity of these scales to the effects of sport-related concussion has been repeatedly demonstrated. Until recently, however, the psychometric properties of these checklists/scales have remained largely unexamined. Piland and colleagues reviewed data from a group of 279 college athletes who were administered a 16-item symptom scale at baseline to explore the factor structure of the scale, which was hypothesized to consist of three relatively distinct domains. They eliminated 7 items, primarily on the basis of face/content validity, and achieved a better fit to their model. Clinical validity was explored with a small sample of concussed players (N=17).

Although this study involved a sophisticated approach to exploring certain psychometric properties of a concussion symptom scale, the primary application of a symptom scale in the medical management of sport-related concussion is in the efficient and sensitive detection of the effects of concussion, as opposed to the characterization of these effects. In this context, item selection should be driven primarily by sensitivity to concussion, requiring an empirical approach to determine item retention. In addition, the Piland et al. study did not explore the scaling characteristics of their instrument, perhaps because of the relatively small number of injured players in their sample. As a result, it remains unclear which symptoms are actually sensitive to the effects of concussion, and whether or not a 7-point Likert-type scale is necessary for the detection and tracking of concussion-related symptomatology.

We recently completed three separate studies, involving the use of largely overlapping symptom scales, with data on over 16,000 athletes at baseline and over 600 athletes following concussion. We combined these datasets in order to enable an empirical study of each item’s value to the scale and to apply Rasch analysis in exploring the scaling characteristics of these symptoms. The purpose of this paper was to derive the most sensitive and efficient scale possible for the detection and tracking of self-reported symptoms following sport-related concussion.

**Methods**

The data for this study were derived from three separate projects: The Concussion Prevention Initiative (CPI), the NCAA Concussion Study (NCAA), and the Project Sideline (Sideline). The protocols and subject
inclusion for each of these projects are described below. The symptom scales employed in each project are contained in Table 1. Each symptom, in each project, was scored on a 7-point Likert-type scale from 0 (absent) to 6 (severe). While the symptom scales for each project differed slightly, they did have substantial overlap with one another, and with symptom scales used in earlier studies.\textsuperscript{27, 28}

**Concussion Prevention Initiative (CPI):** This project involved the collection of prospective data from 14 colleges and 110 high schools from 2000-2003, involving athletes from football, men’s soccer, women’s soccer, men’s lacrosse, women’s lacrosse, men’s ice hockey, and women’s ice hockey. The total number of athletes examined at baseline was 9,094 (72.7% male), with 375 subsequent concussions.

**NCAA Concussion Study (NCAA):** This study involved 4238 male football players from 15 US colleges. All players underwent preseason baseline testing in 1999, 2000, and 2001. There were 196 subsequent concussions, with assessments points at the time of injury, 3 hours post-injury, and at 1, 2, 3, 5, 7, and 90 days post-injury. Portions of the data from this study have been reported elsewhere.\textsuperscript{8, 29}

**Project Sideline (Sideline):** This Milwaukee-based project began in 2000 and involved a total of 18 high schools in the southeastern Wisconsin area, including athletes from football, hockey, and soccer teams. The baseline sample included a total of 3,018 athletes (97% male), with a total of 70 subsequent concussions. Portions of these data have been presented elsewhere.\textsuperscript{43}

Insert Table 1 about here

---

**Analyses/Results**

**Effect size/sensitivity:** The common assessment points for all three studies were immediately post injury, post-game (approximate 3 hours post-injury), day 1, day 3, and day 5. The first analysis was designed to eliminate any items that proved to be insensitive to concussion. The criterion for retention was an effect size of at least .3 on at least 2 of the 5 post-injury assessment points. Although this effect size is modest, we felt that requiring this effect size to be reached on at least two assessment points was an adequately conservative approach to item retention. This resulted in the elimination of 13 variables, leaving the following 14 variables: *Headache, nausea, balance/dizziness, fatigue, drowsiness, feeling slowed down, in a fog, difficulty concentrating, difficulty
remembering, neck pain, blurred vision, sensitivity to light, sensitivity to noise, and sensitivity to light/noise. Because sensitivity to light and sensitivity to noise were independently sensitive, and these variables were combined in only the NCAA dataset, sensitivity to light/noise was also eliminated as a separate variable. In addition, it seemed likely that neck pain was attributable to cervical strain and not a direct result of concussion; as a result, this variable was eliminated as well, leaving a total of 12 symptoms.

Scaling metrics: Rasch analysis was then used to further refine the scale. In Rasch analysis, person “ability” (in this case, athletes reporting the least number of symptoms to those reporting the most) and item “difficulty” or endorsibility (from the most commonly to least endorsed symptom) are calibrated onto a common underlying scale, which is measured in logits, i.e., log odds units. Hence, the probability of an athlete endorsing a symptom on the scale is a logistic function of the relative distance between the number of symptoms endorsed by an athlete and how common the symptom is endorsed. In other words, a highly symptomatic athlete always has a higher probability of endorsing a scale item than a less symptomatic patient. In addition, a rarely endorsed item always has a lower probability of endorsement than a frequently reported item, regardless of how symptomatic an athlete may be. Taken together, a highly symptomatic athlete will have a higher probability of responding “yes” to a rare symptom than a less symptomatic athlete. Rating scale analysis, one of the models within the Rasch measurement family, was used because the items were originally based on a seven-point Likert scale. There were several indicators suggesting that there was insufficient information in the data to yield reliable parameter estimates if a seven-point scale were used, e.g., (a) a number of rating categories had less than 10 observations; (b) irregularity in observation frequency across categories was found that signaled aberrant category endorsement by subjects; (c) average measures did not advance monotonically with category; and (d) some step categories advanced by less than 1.4 logits while others advanced by more than 5.0 logits. These findings indicated that the number of categories needed to be reduced, and these were systematically reduced until optimization was achieved. A dichotomous scale (0=symptom absent, 1=symptom present, at any severity) was determined to be optimal.
The dichotomous Rasch model was then re-fitted. Figure 1 shows the person “ability” (i.e., level of symptomatology) and item “difficulty” (i.e., frequency of item endorsement). Both persons and items are expressed in logits, i.e., logarithmically transformed probabilities of symptom endorsement given a specific level of symptomatology, ranging from about –5.0 to +3.0 logits for the CAS and forming a hierarchical linear scale with equal intervals. The logit measure appears on the left of the figure. Persons are on the left and items are on the right. Items at the bottom of the figure were frequently endorsed while items at the top were rarely endorsed. A person appearing at 0 logits would have a 50% chance of reporting concentration difficulties and feeling like “in a fog.” That same person would have less than a 50% chance of reporting nausea, sensitivity to light and noise, and blurred vision. Conversing, this person would have a greater than 50% chance of reporting drowsiness, fatigue, balance problems, headache, and feeling slowed down.

Relatively good item fit was obtained: average OutFit Mean Square was 0.99 (z = -.4) and there were no negative point-biserial correlations. The item reliability index of .98 was good with item separation of 7.58, indicating adequately dispersed items on the CAS. The item reliability index indicates the replicability of item placements along the pathway if the same CAS items were given to another sample of concussed athletes with comparable symptomatology.

The “person map” appears on the left side of the figure. As can be seen in Figure 1, most subjects did not endorse many symptoms, with many participants clustered at the bottom of the figure and this was reflected in the person reliability index of .71. Headache was the most commonly reported symptom, followed by “feeling slowed down,” but there was a large “gap” between those items. The Outfit MS associated with “difficulty concentrating” suggested that it might be redundant with “feeling like in a fog” (z = 3.1), but we retained this item due to its relatively high endorsement rate. The Outfit MS associated with “difficulty remembering” indicated a lack of unidimensionality with the other items (z = 2.9). This makes a certain degree of sense, as this item is the most purely cognitive symptom, and we retained it because it was sensitive and filled the gap between the “concentration” and “nausea” items.

**Receiver Operating Characteristics (ROC) curves:** To ensure that we did not lose substantial sensitivity by changing the item scaling from a 7-point Likert scale to a 2-point dichotomous scale, we conducted ROC analyses of data from two assessment points: Immediately post-injury and Day 5 post-injury. Scores for all concussed
players on each scaling approach were compared to the scores for the entire baseline sample. As can be seen in Figure 2, there is almost complete overlap of the ROC curves for the two scaling approaches, with minimal difference in sensitivity or specificity observable between the Likert and dichotomous scaling.

Concussion Symptom Inventory (CSI)

The newly-derived Concussion Symptom Inventory is presented in Appendix 1. To our knowledge, this is the first scale that has been empirically-derived for the purpose of monitoring subjective symptoms following concussion. The source data also constitute the largest sample of prospectively studied cases of concussion in the extant literature, with a concussed sample size of 641 athletes compared a baseline sample of 16,350 athletes. We have elected not to present baseline normative data or attempt to derive “cutoff” scores, because we lack sufficient empirical data at this point to suggest that there actually is a quantifiable risk of returning a player to competition based upon a particular CSI score. This is, of course, true of any symptom scale or other technique for measuring impairment following concussion.

We propose that athletic trainers and team medical personnel employ the CSI as a standardized methodology for tracking symptom resolution following concussion, and incorporate the information from the CSI into clinical decision-making regarding return to play. This decision-making process should be informed by the evolving literature on the natural history and outcome of sport-related concussion, and by the specific clinical circumstances of the individual player. It is important to emphasize that the CSI is not intended to constitute the sole basis for clinical decision-making in the medical management of sport-related concussion, and that individual players may also experience concussion-related symptoms (e.g., sleep disturbance) that are not recorded within the CSI due to the relative infrequency with which they occurred in our concussed sample.

The CSI does, however, provide an empirically-based, relatively rapid, and systematic methodology for tracking subjective symptoms following sport-related concussion. The risks of “premature” return to play following sport-related concussion are as yet poorly-delineated, and none of the many guidelines that have been promulgated for this purpose are evidence-based. They are all in agreement, however, that players should be
symptom-free before being cleared to return. This would seem to be a reasonably conservative approach to concussion management, particularly in younger athletes, until additional data regarding risks are accrued and clinical decision-making can be driven by reliable evidence.
References


Table 1. Symptoms included in scales for each of the three projects

<table>
<thead>
<tr>
<th>Sideline</th>
<th>NCAA</th>
<th>CPI</th>
</tr>
</thead>
<tbody>
<tr>
<td>HEADACHE</td>
<td>HEADACHE</td>
<td>HEADACHE</td>
</tr>
<tr>
<td>NAUSEA</td>
<td>NAUSEA</td>
<td>NAUSEA</td>
</tr>
<tr>
<td>VOMITING</td>
<td>VOMITING</td>
<td>VOMITING</td>
</tr>
<tr>
<td>BALANCE PROBLEMS/DIZZINESS</td>
<td>BALANCE PROBLEMS/DIZZINESS</td>
<td>BALANCE PROBLEMS/DIZZINESS</td>
</tr>
<tr>
<td>FATIGUE</td>
<td>FATIGUE</td>
<td>FATIGUE</td>
</tr>
<tr>
<td>TROUBLE SLEEPING</td>
<td>TROUBLE FALLING ASLEEP</td>
<td>TROUBLE SLEEPING</td>
</tr>
<tr>
<td>SLEEPING MORE THAN USUAL</td>
<td>SLEEPING MORE THAN USUAL</td>
<td>SLEEPING MORE THAN USUAL</td>
</tr>
<tr>
<td>DROWSINESS</td>
<td>DROWSINESS</td>
<td>DROWSINESS</td>
</tr>
<tr>
<td>SADNESS</td>
<td>SADNESS</td>
<td>SADNESS</td>
</tr>
<tr>
<td>NUMBNESS/TINGLING</td>
<td>NUMBNESS/TINGLING</td>
<td>NUMBNESS/TINGLING</td>
</tr>
<tr>
<td>FEELING LIKE “IN A FOG”</td>
<td>FEELING LIKE “IN A FOG”</td>
<td>FEELING LIKE “IN A FOG”</td>
</tr>
<tr>
<td>DIFFICULTY CONCENTRATING</td>
<td>DIFFICULTY CONCENTRATING</td>
<td>DIFFICULTY CONCENTRATING</td>
</tr>
<tr>
<td>DIFFICULTY REMEMBERING</td>
<td>DIFFICULTY REMEMBERING</td>
<td>DIFFICULTY REMEMBERING</td>
</tr>
<tr>
<td>SENSITIVITY TO LIGHT</td>
<td>SENSITIVITY TO LIGHT</td>
<td>SENSITIVITY TO LIGHT</td>
</tr>
<tr>
<td>BLURRED VISION</td>
<td>BLURRED VISION</td>
<td>BLURRED VISION</td>
</tr>
<tr>
<td>SENSITIVITY TO NOISE</td>
<td>SENSITIVITY TO NOISE</td>
<td>SENSITIVITY TO NOISE</td>
</tr>
<tr>
<td>FEELING SLOWED DOWN</td>
<td>FEELING SLOWED DOWN</td>
<td>FEELING SLOWED DOWN</td>
</tr>
<tr>
<td>IRRITABILITY</td>
<td>IRRITABILITY</td>
<td>IRRITABILITY</td>
</tr>
<tr>
<td>NERVOUSNESS</td>
<td>NERVOUSNESS</td>
<td>NERVOUSNESS</td>
</tr>
<tr>
<td>SENSITIVITY TO LIGHT/NOISE</td>
<td></td>
<td>SENSITIVITY TO LIGHT/NOISE</td>
</tr>
<tr>
<td>NECK PAIN</td>
<td>NECK PAIN</td>
<td>NECK PAIN</td>
</tr>
<tr>
<td>FEELING LIKE “IN A FOG”</td>
<td></td>
<td>FEELING LIKE “IN A FOG”</td>
</tr>
<tr>
<td>SKIN RASH/ITCHING*</td>
<td>SKIN RASH/ITCHING*</td>
<td>SKIN RASH/ITCHING*</td>
</tr>
<tr>
<td>TEETH HURTING*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JOINT STIFFNESS*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BURNING FEELING IN FEET*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Included as tests of valid responding/specificity. For all three studies, symptoms were recorded on a 7-point Likert-type scale, with scores ranging from 0 (absent) to 6 (severe).
Figure 1. Rasch Analysis Person-Item Map of Concussion Symptom Inventory
(see text for explanation)

persons MAP OF items

3

.## +

.T|

. |

.T|

2

.# +

. |

noise

. |

. |

SIS

# light vision

1

+ nausea

.# |

. # |

. # |

. | remember

. |

0

+M concen fog

.##

. | drowsiness

. M

. | balance fatigue

. | slowed

.-1

+.

.# |

. S

.-2

+.

.T|

.-3

.# +

headache

.T|

.-4

.#################

<less><frequ>

EACH # IS 18.
Figure 2: Receiver Operating Curve Analyses

These ROC curves reflect the sensitivity and specificity of the 2-item CAS scale using Likert scaling (0=absent to 6=severe) in comparison to dichotomous scaling (absent or present). The top graph compares controls to concussed athletes immediately post-concussion, and the bottom graph compares controls to concussed athletes 5 days post-concussion. The Likert and dichotomous scaling curves are virtually overlapping at both time points, reflecting nearly identical sensitivity/specificity.
<table>
<thead>
<tr>
<th>Symptom</th>
<th>Absent</th>
<th>Present</th>
</tr>
</thead>
<tbody>
<tr>
<td>HEADACHE</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NAUSEA</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>BALANCE PROBLEMS/DIZZINESS</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>FATIGUE</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>DROWSINESS</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>FEELING LIKE “IN A FOG”</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>DIFFICULTY CONCENTRATING</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>DIFFICULTY REMEMBERING</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SENSITIVITY TO LIGHT</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SENSITIVITY TO NOISE</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>BLURRED VISION</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>FEELING SLOWED DOWN</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

TOTAL ________