An Evaluation of the Clinical Utility of the Sport Concussion Assessment Tool 3 (SCAT3) in Varsity Athletes

By

Rachel Iris Downey

A thesis submitted in conformity with the requirements for the degree of Master of Science
Rehabilitation Science Institute
University of Toronto

© Copyright by Rachel Downey, 2016
An Evaluation of the Clinical Utility of the Sport Concussion Assessment Tool 3 (SCAT3) in Varsity Athletes

Rachel Iris Downey
Master of Science
Rehabilitation Science Institute
University of Toronto
2016

Abstract

Sport-related concussion is a type of traumatic brain injury that remains challenging to identify and diagnose. The development of the Sport Concussion Assessment Tool-3 (SCAT3) provides a strong framework for evaluating concussion by amalgamating well-established assessment tools. The current study evaluated the clinical utility of the SCAT3 in identifying concussion across time by comparing varsity athletes’ post-concussion scores to pre-injury baseline scores, and also to normative data. Results suggest that the SCAT3 is sensitive to detecting concussion within 3 to 5 days and 3 weeks post-injury, with the utility being greatest using a normative approach, compared to the baseline-retest method. The symptom scale included in the SCAT3 largely contributed to the overall sensitivity of the tool, followed by the balance and cognitive measures. This study provides evidence to support the use of the SCAT3 within the first 3 weeks post-injury, however might be limited beyond this timeframe.
# Table of Contents

Table of Contents ........................................................................................................ iii

List of Tables .................................................................................................................. vii

List of Figures ................................................................................................................ ix

List of Appendices ......................................................................................................... xi

Chapter 1 Introduction .................................................................................................. 1

Chapter 2 Literature Review ......................................................................................... 3

2.1 Concussion: Defined ............................................................................................... 3

2.2 Concussion incidence and risk ............................................................................ 5

2.3 Biomechanics of concussion ............................................................................. 6

2.4 Pathophysiology of concussion ......................................................................... 6

2.5 Concussion assessment: ................................................................................... 8

2.5.1 Baseline testing ............................................................................................... 8

2.5.2 Neuropsychological testing .......................................................................... 9

2.5.5 Symptom evaluation ..................................................................................... 10

2.5.4 Physical examination .................................................................................. 11

2.5.5 Sideline examination ................................................................................... 12

2.5.6 The Sport Concussion Assessment Tool ................................................... 12

2.6 Test validity ......................................................................................................... 14

2.7 Modifying factors to concussion testing ........................................................... 14

2.7.1 Age .............................................................................................................. 14

2.7.2 Sex ............................................................................................................. 15

2.7.3 Learning disability, psychiatric condition, history of concussion ............. 15

2.8 Clinical management of concussion ................................................................ 15

2.9 Rationale for current research .......................................................................... 16
Chapter 3 normative SCAT3 values and modifying factors in varsity athletes playing a high-risk sport

3.1 Introduction

3.2 Rationale

3.3 Research objectives

3.4 Methods and procedure

3.4.1 Design

3.4.2 Participants

3.4.3 Outcome measures

3.4.4 Procedure and data collection

3.4.5 Data analysis

3.5 Results

3.5.1 Demographics

3.5.2 Modifying factors

3.5.3 Covariates

3.6 Discussion

3.7 Conclusions

Chapter 4 Clinical utility of the SCAT3

4.1 Introduction

4.2 Rationale
5.2.1 SCAT3..............................................................................................................101

5.2.2 ANAM............................................................................................................102

5.2.3 Baseline versus normative data........................................................................102

5.2.4 Future research directions.................................................................................103

5.2.5 Clinical recommendations..................................................................................104

Chapter 6 Conclusions and implications....................................................................107

References....................................................................................................................108

Appendices..................................................................................................................114
# List of Tables

| Table 3.1 | Number of participants per sport | 25 |
| Table 3.2 | Number of participants per year of education | 25 |
| Table 4.1 | Pearson correlation coefficient values for SCAT3 scores | 69 |
| Table 4.2 | Sensitivity and specificity of SCAT3 components individually and combined at the acute phase of injury using a test-retest and normative approach | 70 |
| Table 4.3 | Sensitivity and specificity of SCAT3 components individually and combined 3 weeks post-injury using a test-retest and normative approach | 72 |
| Table 4.4 | SCAT3 sensitivity ranking by methodology at acute phase of injury | 75 |
| Table 4.5 | SCAT3 sensitivity ranking by methodology at 3 weeks post-injury | 76 |
| Table 4.6 | Pearson correlation coefficient values for ANAM throughput scores | 78 |
| Table 4.7 | Pearson correlation coefficient values for ANAM reaction time scores | 78 |
| Table 4.8 | Pearson correlation coefficient values for ANAM percent correct scores | 79 |
| Table 4.9 | Sensitivity and specificity of ANAM component throughput scores individually and combined 3 to 5 days post-injury using a test-retest and normative approach | 80 |
| Table 4.10 | Sensitivity and specificity of ANAM component percent correct scores individually and combined 3 to 5 days post-injury using a test-retest and normative approach | 80 |
Table 4.11  Sensitivity and specificity of ANAM component reaction time scores individually and combined 3 to 5 days post-injury using a test-retest and normative approach.......................................................................................................................................81

Table 4.12  Sensitivity and specificity of ANAM component throughput scores individually and combined 3 weeks post-injury using a test-retest and normative approach.......................................................................................................................................82

Table 4.13  Sensitivity and specificity of ANAM component percent correct scores individually and combined 3 weeks post-injury using a test-retest and normative approach.......................................................................................................................................82

Table 4.14  Sensitivity and specificity of ANAM component reaction time scores individually and combined 3 weeks post-injury using a test-retest and normative approach.......................................................................................................................................83

Table 4.15  Sensitivity and specificity of combined throughput, percent correct, and reaction time scores for each ANAM subtest at acute phase of injury..............................................84

Table 4.16  Sensitivity and specificity of combined throughput, percent correct, and reaction time scores for each ANAM subtest at 3 weeks post-injury..............................................84

Table 4.17  ANAM sensitivity ranking by methodology at acute phase of injury.........................85

Table 4.18  ANAM sensitivity ranking by methodology 3 weeks post-injury.............................86
List of Figures

Figure 3.1  Median PCSS total symptom score greater in males compared to females........26
Figure 3.2  Median PCSS symptom severity score greater in males compared to females.....26
Figure 3.3  Mean ranking of SAC performance greater in females compared to males........27
Figure 3.4  Mean ranking of PCSS total symptoms greater in athletes with a history of mental health compared to those without. .................................................................28
Figure 3.5  Mean ranking of PCSS symptom severity greater in athletes with a history of mental health compared to those without.................................................................29
Figure 3.6  Mean ranking of PCSS total symptoms greater in athletes with a familial mental health history compared to those without.................................................................30
Figure 3.7  Mean ranking of PCSS total symptoms greater in athletes with a family history of mental health compared to those without.................................................................30
Figure 3.8  Median SAC score lower in athletes with a history of a learning disability or ADHD compared to athletes without..................................................................................31
Figure 4.1  Visual representation of testing timeline.........................................................52
Figure 4.2  Mean ranking of PCSS total symptom score significantly greater in concussed athletes at acute phase of injury compared to controls..............................................58
Figure 4.3  Mean ranking of PCSS symptom severity score significantly greater in concussed athletes at acute phase of injury compared to controls..............................................58
Figure 4.4  Main effect of time within control athletes on PCSS total symptom score........60

Figure 4.5  Main effect of time within control athletes on PCSS symptom severity score.....60

Figure 4.6  Mean ranking of PCSS total symptom score in concussed athletes significantly
greater at acute phase of injury compared to baseline and 3 weeks post-injury....62

Figure 4.7  Mean ranking of PCSS symptom severity score in concussed athletes significantly
greater at acute phase of injury compared to 3 weeks post-injury......................62

Figure 4.8  Baseline m-BESS ROC curve significantly greater than female normative
approach............................................................................................................89

Figure 4.9  Significantly lower SAC AUC compared to total symptom and symptom severity
score at acute phase of injury in male normative data.................................90

Figure 4.10 Significantly greater total symptom AUC compared to symptom severity, SAC,
and m-BESS at acute phase of injury in entire normative dataset.......................91
List of Appendices

Appendix 1  Information sheet/consent form..............................................................114

Appendix 2  Copy of SCAT3 form with major components: PCSS, SAC, m-BESS...........118
Chapter 1
Introduction

Sport-related concussion has emerged as a major public health concern following two decades of mounting evidence suggesting its potential consequences involving a myriad of physical, cognitive, emotional, psychiatric, and sleep disturbances. The incidence of sport-related concussions in the United States is an estimated 1.6 to 3.8 million, with a large proportion of concussions occurring in the varsity athletic context (Daneshvar, Nowinski, McKee, & Cantu, 2011).

Research revealing the risk associated with recurrent concussions and second-impact syndrome precipitated the development of sideline evaluation tools to quickly and accurately identify when a concussion is sustained (McCrea, 2001). The Sport Concussion Assessment Tool (SCAT) was first established as a sideline measure to evaluate concussion following the Second International Symposium on Concussion in Sport in 2004 and has since been revised to its current version, the SCAT3 (McCrory et al., 2005, 2013). The SCAT3 provides a strong framework for evaluating concussion both on the sideline and in the clinic by amalgamating well-established neurological signs and self-reported symptoms. Although experts in the field have recommended widespread use of SCAT3 (McCrory et al., 2013), the validity, and hence the clinical utility, of this measure in identifying concussion over time has not been explored empirically.

The purpose of this study was therefore to examine the clinical utility of the Sport Concussion Assessment Tool-3 (SCAT3) in a cohort of varsity athletes. Primarily, this involved investigating the sensitivity and specificity of the SCAT3 component scores independently and as a global measure across time to determine when and which measures within the SCAT3 are most optimal in identifying sport related concussion (SRC). The sensitivity and specificity of the SCAT3 was calculated using two different approaches in order to determine how the utility of this measure differs in the absence or presence of an athlete’s baseline scores. Firstly, a test-retest approach was implemented, where an athlete’s post-injury scores were compared to his or her own pre-injury baseline scores, and secondly a normative approach was used, where athletes’ post-injury scores were compared to a database of scores taken from varsity athletes at the University of Toronto. The sensitivity and specificity of a common computerized neurocognitive measure, the
Automated Neuropsychological Assessment Metrics (ANAM), was also examined in order to compare it against the utility of the SCAT3 in identifying SRC. A secondary objective of this study was to establish normative baseline data for the major SCAT3 components in a Canadian varsity athletic sample and to identify any factors that may potentially modify baseline scores.

Chapter 2 of this thesis provides a review of the literature that guided the rationale for the current study. Chapters 3 and 4 are structured as independent manuscripts that examine the separate research objectives. Chapter 5 provides a general discussion of the thesis in its entirety, along with future research directions and clinical recommendations.
Chapter 2

Literature Review

2.1 Concussion: Defined

One of the difficulties with concussion identification and management lies in the ever-evolving definition and terminology used to characterize the injury itself. Indeed, the term ‘concussion’ is thought to date back as early as 460-360 B.C., with transcriptions of Hippocrates reporting on the clinical symptoms present after head injury (McCrory & Berkovic, 2001). As advancements in the field have generated revisions to the definition of concussion, it is critical that a standardized and modernized definition be utilized to guide future research and clinical practice. According to the consensus statement taken from the 4th International Conference on Concussion in Sport, experts acknowledge concussion as a type of traumatic brain injury (TBI), defined as “a complex pathophysiological process affecting the brain, induced by traumatic biomechanical forces” (McCrory et al., 2013).

The consensus statement highlighted several defining features of concussion to include: 1) that a concussion may be caused by either a direct or indirect blow to the head, 2) that symptoms resulting from concussion are typically of rapid onset, are transient, and resolve spontaneously, however in some cases may take time to evolve, 3) that functional disturbances in the acute phase of concussion are represented by neuropathological changes, but lack structural changes as observed by standard neuroimaging techniques, and 4) that concussion results in a graded set of symptoms that may or may not involve a loss of consciousness (LOC). Moreover, the trajectory of recovery following concussion is typically sequential, but may be prolonged in some cases (McCrory et al., 2013).

While although this definition of concussion is now widely accepted, it is not directly comparable to how a concussion may be defined in clinical practice. Specifically, as current imaging modalities and other tools used to evaluate concussion lack direct prognostic validity, the clinician depends on indirect measures for evaluating brain function to determine injury severity and recovery (Makdissi, 2009). The signs and symptoms resulting from concussion can be broadly categorized into four main domains: somatic (including headache or balance problems), cognitive (such as difficulty concentrating), emotional (including irritability or
sadness), and sleep disturbance (such as trouble falling asleep or sleeping more than usual) (Graham, Rivara, Ford, & Spicer 2014).

Historically, clinicians have attempted to classify the severity of a concussive injury by the use of grading scales. Several grading scales exist, including those established by Cantu (Cantu, 1986; 2001), the Colorado Medical Society (CMS, 1990), and the American Academy of Neurology (AAN, 1997). Such scales were developed to categorize concussion as either a grade 1, 2, or 3 injury, with ‘grade 3’ denoting the greatest severity, according to parameters of amnesia and LOC. However, noting that any such grading scales essentially lack any validity, the trend in recent years has moved away from concussion grading scales and towards a more thorough neurological assessment to guide return-to-play decisions (Giza, Kutcher, Ashwal, Barth, Getchius, et al., 2013).

Current clinical practice is that an athlete not return to sport participation until the signs and symptoms of concussion are completely recovered without symptom exacerbation following physical exertion (McCrory et al., 2013). Research has demonstrated that in 80-90% of individuals with concussion, signs and symptoms resolve on average within 7 to 10 days (Makdissi, 2010; McCrory et al., 2013). However, a small proportion of individuals may experience prolonged symptoms, with less than 1% of individuals being diagnosed with post concussion syndrome as characterized by the experience of persistent symptoms for greater than three months (Makdissi, 2010; Ryan & Warden, 2003).

Further adding confusion to the classification of concussion was the introduction of the concepts of “simple” and “complex” concussion (McCrory et al., 2005). A simple concussion was defined as the most common form of injury, characterized by the resolution of symptoms within the typical 7 to 10 day timeframe, with the athlete resuming sport participation with no complications (Makdissi, 2009). In contrast, a complex concussion was hypothesized to have occurred when athletes suffered persisting symptoms either at rest or following exertion, prolonged LOC, or prolonged cognitive impairment (Makdissi, 2009). However, the simple versus complex designation could only be established at outcome and it therefore lacked prognostic value. Accordingly, the differential terms ‘simple’ and ‘complex’ have been
abandoned by the scientific community, with the recommendation for concussions to be considered as a “single entity” that may be susceptible to modifying factors (Makdissi, 2009). The definition and overall knowledge surrounding concussion as a unique neurological injury has evolved over the last 20 years, thereby allowing for better clinical management of athletes and others. However, given that the terminology used to identify and classify concussion is not well established, the true epidemiology of concussion is difficult to ascertain (King, Brughelli, Hume, & Gissane, 2014).

2.2 Concussion incidence and risk

It is now well known that physical activity is important in supporting an individual’s health, personal well-being, and overall quality of life. Research consistently demonstrates that individuals who are physically active live longer lives and have a lower risk for developing health complications, such as Type II diabetes and heart disease (Reiner, Niermann, Jekauc, & Woll, 2013). The role of physical exercise in sport has similarly proven effective in maintaining mental health by encompassing a social aspect to reduce stress, allow for personal growth, and enhance well-being (Eime, Young, Harvey, Charity, & Payne, 2013).

Although there are sizeable benefits of sport participation, there is also an increased risk of injury. Concussion, in particular, is a common injury across all sports, and is especially prevalent in contact sports (Okonkwo, Tempel, & Maroon, 2014). The incidence of concussion in sport in the United States is estimated to be within the range of 1.6-3.8 million (Langlois, Rutland-Brown, & Wald, 2006). This broad and somewhat ambiguous rate is an extrapolation based on an outdated study by the Centers for Disease Control and Prevention (CDC) estimating the incidence of concussion with LOC to be 300,000, while considering that only 8-19% of individuals who have sustained a concussion experience any loss of consciousness (Thurman, Branche, & Sniezek, 1998).

This rate is suggested to be a vast underestimate of the true epidemiological incidence of concussion due to the small number of athletes who seek medical attention, however is still considered to have risen substantially within the last decade (Daneshvar et al., 2011). Proposed explanations for this increase have included stronger and faster impacts without the complete
effectiveness of protective headgear, in addition to the heightened awareness and education surrounding the importance of recognizing concussion as a serious injury (Daneshvar et al., 2011). As current data available on the CDC does not provide statistics specific to the rate of SRC, the true incidence remains unknown.

Epidemiological research, however, has allowed for a better understanding of specific risks associated with sustaining a concussion in sport. Such findings outline that concussions are of highest risk in American football and ice hockey, in females compared to males, during competition as compared to practice, in younger athletes, in individuals with a history of pre-existing migraines, and in individuals having a previous history of one or more concussions (Daneshvar et al., 2011; Harmon et al., 2013; Koh, Cassidy, & Watkinson, 2003). Much of the evidence describing these physiological risk factors is grounded in basic biomechanical research.

### 2.3 Biomechanics of concussion

Measurement of the biomechanical forces underlying concussion has been investigated thoroughly. Research findings utilizing animal models, cadavers, surrogates, and computer models suggest the role of linear and rotational head acceleration as a primary mechanism of concussion, with recent research emphasizing the significance of rotational shearing specifically in leading to the diffuse symptoms characteristic of concussion (Guskiewicz & Mihalik, 2011; Meaney & Douglas, 2011; Poirier, 2003).

Concussions may further be classified as a coup or contra-coup injury using this biomechanical framework. Coup injuries are commonly produced when a moving force is applied to the head at rest, resulting in a head injury on the same side of contact (Poirier, 2003). In comparison, contra-coup injuries are typically the outcome of a moving head striking an object that is at rest, resulting in head injury on the opposite site of contact (Poirier, 2003).

Concussions may also be categorized differentially as either contact or inertial head injuries, where the source of head acceleration is caused by either a direct or indirect strike, respectively (Meaney & Douglas, 2011). More recent research has investigated other important
biomechanical variables that may elicit concussion, including force, jerk, impulse, and impact duration (Broglio et al., 2009).

### 2.4 Pathophysiology of concussion

In line with investigations into the physical forces underlying concussion is research examining the pathophysiological disturbances that occur following this type of injury. In consequence of either linear or rotational head acceleration, common neuropathophysiological sequelae are thought to encompass neurometabolic, neurochemical, and cerebrovascular regulation changes (Makdissi, 2009; Tan, Meehan, Iverson, & Taylor, 2014).

Data derived from animal models using experimental head injury paradigms, as well as in human head injury patients using techniques such as cerebral microdialysis, positron emission topography (PET), and magnetic resonant spectroscopy (MRS), suggest that following concussion, a series of neurometabolic and neurochemical events are initiated (Giza & DiFiori, 2011). Specifically, the excitatory neurotransmitter, glutamate, is released into extracellular space, causing mass depolarization of neurons and disruption of homeostasis amongst intra- and extracellular sodium and potassium ions (Giza & DiFiori, 2011). In order to restore this imbalance, adenosine triphosphate (ATP) is required as an energy source to fuel membrane ionic pumps, resulting in increased cerebral glucose oxidation (Giza & DiFiori, 2011). As a consequence of this metabolic demand, cerebral blood flow may be disrupted, making cells susceptible to long-term damage (Giza & DiFiori, 2011; Tan et al., 2014). The role of glutamate also induces neurotoxicity by binding to N-methyl-d-aspartate (NMDA) receptors, causing an influx of calcium ions (Signoretti, Lazzarino, Tavazzi, & Vagnozzi, 2011). Overloading of intracellular calcium may cause mitochondrial dysfunction, leading to subsequent oxidative stress, and the potential for apoptosis, or programmed cell death (Giza & DiFiori, 2011; Signoretti et al., 2011).

There are inconsistent findings in correlating such neurometabolic and neurochemical changes with clinical symptoms and recovery. Specifically, research in rats shows resolution from this metabolic crisis within 7 to 10 days following injury, a trajectory consistent with typical recovery time (Giza & Hovda, 2011). However, in humans the neurometabolic cascade is
suggested to recover between 2 and 4 weeks post-injury, suggesting that there might be a
decoupling of overt clinical symptoms and the underlying pathophysiology of concussion (Giza
& DiFiori, 2011; Giza & Hovda, 2015; Giza & Hovda, 2001). As such, a definitive link
between neurobiological changes and concussive symptoms remains absent.

2.5 Concussion assessment:

2.5.1 Baseline testing

Although early detection is the mainstay for managing SRC, clinical evaluation of the injury
remains challenging (Okonkow Tempel, & Maroon, 2014). Pivotal to the progress achieved in
concussion assessment has been the implementation of baseline testing. In the athletic context,
baseline testing involves the assessment of an athlete’s level of functioning when the athlete is
healthy, typically during preseason training. In the case where a concussion is sustained, the
athlete can be re-tested and the results can be compared to his or her own pre-injury baseline
scores. As variability is inherent in concussion, this methodology allows for individualized
assessment that may be useful in guiding clinical decisions (Slobounov & Sebastienelli, 2006).
Baseline testing has also allowed for the identification of potential modifying factors that may
affect test interpretation. Such factors indicated in modifying baseline concussion assessments
include age, gender, test setting, fitness level, and a history of a psychiatric condition or a
learning disability (Dessy et al., 2014).

Although the baseline-post injury paradigm may be useful in deterring against potential errors
derived from a normative approach given these potential modifiers, baseline testing also has
inherent limitations. One conceivable issue with baseline testing is the athlete’s level of
motivation and its influence on test scores (Bailey, Echemendia, & Arnett, 2006). Specifically, if
an athlete is motivated to return-to-play quickly in the case that they sustain a concussion, they
may purposely perform poorly on cognitive and balance measures and maximize symptoms at
baseline (Bailey, Echemendia, & Arnett, 2006). Another potential issue with baseline testing is
grounded in the psychometric properties inherent to retesting, with the possibility of practice
effects leading to improved scores following neurocognitive testing that may affect post-
concussion test interpretation (Ragan & Kang, 2007; Randolph, McCrea, & Barr, 2005). Lastly,
baseline testing requires a large time and financial commitment that may not be feasible for all institutions or communities supervising sport participation.

Although experts do not endorse baseline testing as a mandatory requirement, its implementation may be beneficial in informing test interpretation (McCrory et al., 2013). As such, baseline testing is now widely employed across a range of assessment tools measuring neurocognitive functioning, symptom endorsement, and postural stability.

### 2.5.2 Neuropsychological testing

Barth and colleagues (1989) pioneered the implementation of pre-injury baseline neuropsychological (NP) testing under the Sports as a Laboratory Assessment Model to better characterize the neurocognitive sequelae following concussion in athletes. NP testing batteries have proven to be objective measures for detecting change between pre- and post-concussion functioning in the domains of working memory, attention and concentration, information processing, and reaction time (Schatz & Covassin, 2006).

Historically, paper and pencil tasks have been utilized to measure neurocognitive functioning. These traditional models of NP assessment, however, are not optimal in the athletic context due to the logistical difficulties associated with testing a large population, including the sizeable time and financial commitment to administer and score such extensive measures (Randolph, McCrea, & Barr, 2005). Moreover, the psychometric properties of NP testing may limit serial testing necessary for assessing the neurocognitive functioning of recovery and therefore in making safe return-to-play decisions (Barr & McCrea, 2001). Given such concerns associated with conventional NP testing in the athletic context, in conjunction with technological advancements, computerized NP testing batteries have become increasingly utilized.

Computerized testing batteries offer significant advantages compared to conventional testing methods, including brevity of time to complete the test, more thorough standardization of administration, built-in algorithms to ease scoring, reduced practice effects (by creating alternate forms and randomizing test stimuli), and a more accurate measure of reaction time (Cernich, Reeves, Sun, & Blieberg, 2007). There are now a considerable amount of computerized batteries available for the assessment of concussion that measure similar cognitive domains as
conventional tests, including memory, attention and concentration, information processing speed, and reaction time (Cernich et al., 2007).

The Automated Neuropsychological Assessment Metrics (ANAM) and the Immediate Post Concussion Assessment Metrics (ImPACT) are two commonly used computerized concussion assessment batteries. Both test batteries measure similar cognitive domains including reaction time, processing efficiency, working memory, sustained attention, concentration and spatial processing (ANAM4™, 2010; Schatz, 2010).

Although NP testing may allow for a more objective approach to identifying concussion and tracking its resolution, as compared to more subjective, self-report symptom measures, some research demonstrates discordance between symptom and cognitive resolution (McCrorry et al., 2013). It is common for cognitive recovery to follow symptom resolution, with specific findings demonstrating that a moderate percentage of athletes (38% in a sample of 21 concussed varsity athletes) continue to present with deficits in some cognitive domains following symptom resolution (Broglio, Macciocchi, & Ferrara, 2007). Therefore, a multifaceted approach is warranted in order to ensure that an athlete’s clinical symptoms and cognitive deficits are completely resolved prior to returning to play (Collie, Makdissi, Maruff, Bennell, & McCrorry, 2006).

2.5.5 Symptom evaluation

Evaluation of concussion-related symptoms is essential for informing concussion identification and management. However, a major issue with symptom assessment lies within its subjective nature, providing the athlete the opportunity to underreport or frankly deny any symptoms in order to return to play. As noted, an athlete’s level of motivation may obscure test interpretation if an athlete minimizes the actual extent of his or her symptoms in order to continue sport participation (Schatz & Covassin, 2006). Furthermore, many of the items included in symptom measures are non-specific to concussion and may overlap with other co-morbidities, such as anxiety and depression, or lifestyle factors, including physical exercise and level of fatigue (Balasundaram, Athens, Schneiders, Mccrorry, & Sullivan, 2016). Despite these limitations, symptom reporting is an important component of concussion assessment and management.
Symptom evaluation tools include checklists and/or graded symptom scales. Checklists include known concussion-related symptoms where an athlete can respond either “yes” in the case that they are experiencing the symptom, or “no” to indicate that the athlete is not experiencing the symptom (Guskiewicz et al., 2004). Symptom scales may include similar questions as a symptom checklist, however it provides a summative severity index of the symptoms being experienced, commonly incorporating a Likert scale that ranges from 0 (to denote no symptoms) to 6 (indicating severe symptoms) (Guskiewicz et al., 2004).

Existing symptom scales include the Graded Symptom Checklist (GSC), the Post Concussion Scale (PCS), and the Post Concussion Symptom Scale (PCSS). Although the GSC is termed a “checklist”, it has since been revised to include a severity index that utilizes a Likert scale, making it more akin to a symptom scale (Valovich Mcleod & Leach, 2012). The use of symptom scales are recommended clinically as they provide a more detailed assessment of symptom variability while allowing for the detection of more subtle changes, as compared to the dichotomous scale inherent to symptom checklists (Valovich Mcleod & Leach, 2012). Because several iterations of symptom scales exist, with slight variations in the items included in the evaluations, empirical between-scales comparisons are difficult.

### 2.5.4 Physical examination

In addition to cognitive changes and increased somatic symptoms, impairments in motor functioning, such as balance and postural stability, are also common following concussion. Such impairments may have fundamental consequences for the athlete as it may lead to an increased risk of injury if returned to play prematurely, prior to complete recovery (McCrory et al., 2013). The primary tools used to assess for balance function and postural stability following concussion include the Balance Error Scoring System (BESS) and the Sensory Organization Test (SOT). The BESS assesses balance control across three different leg stances while the athlete stands on either a hard or soft testing surface (SCAT3, 2013). A recent systematic review proposed the BESS as a valid tool for detecting large deficits following concussion, however the validity of this tool was questioned when deficits appear more subtle (Bell, Guskiewicz, Clark, & Padua, 2011).
The SOT is a measure of postural control that specifically assesses visual, somatosensory, and vestibular sensory functioning (Register-Mihalik et al., 2013). Although research has found sensory interaction and balance deficits using the SOT after concussion, these typically resolve within 3 to 5 days following injury (Guskiewicz, 2011). Additionally, as the SOT utilizes a force platform to measure these balance domains, its applicability and practicality for clinicians has come into question (Guskiewicz, 2011).

2.5.5 Sideline examination

Sideline evaluation is of paramount importance in determining immediate management of an athlete suspected of sustaining a concussion and to rule out more serious brain injury (Guskiewicz & Broglio, 2011). In the varsity context, the team’s athletic therapist commonly conducts the sideline assessment during either a practice or game when an athlete is suspected of sustaining a concussion.

Concussion assessments utilized on the sideline are brief in nature and examine typical impairments and deficits associated with concussion. Common tools include the Glasgow Coma Scale (GCS), the Standardized Assessment of Concussion (SAC), the BESS, and symptom grading scales. The GCS may be used to categorize the severity of brain injury, by measuring level of consciousness via eye, motor, and physical response (Graham et al., 2014). As concussions rarely result in prolonged loss of consciousness or absence of responsivity, the GCS score is typically rated between 13 to 15, with lower scores designating greater injury severity (Bodin, Yeates, & Klamar, 2012).

The SAC is a brief cognitive measure used to identify impairments in the domains of orientation, concentration, and memory (Okonkwo et al., 2014). The use of symptom scales and the BESS on the sideline can similarly detect acute impairments that can inform clinical management. As stand-alone measures are often limited in informing the diagnosis of concussion, a multifaceted approach on the sideline is necessary.
2.5.6 The Sport Concussion Assessment Tool

The Sport Concussion Assessment Tool (SCAT) is a sideline tool that was first introduced by expert consensus following the 2nd International Conference on Concussion in Sport in 2004 and was developed by combining the best available sideline concussion assessment tools in use at the time (Guskiewicz et al., 2013; McCrory et al., 2005). Such measures integrated into the SCAT included a brief neurological screen, a graded symptom scale, the SAC, and a Maddocks score which included questions regarding the athlete’s orientation specific to the sport context (McCrory et al., 2005).

The original version of SCAT was then revised as the SCAT2 following the 3rd International Conference on Concussion in Sport in 2008 (McCrory et al., 2009). Modifications to the tool included the addition of the GCS, as well as an objective balance measure (a modified version of the BESS to only include a hard testing surface) (McCrory et al., 2009). Additionally, changes were made to the items on the symptom scale, the use of the SAC, and the production of a total score (Zimmer, Marcinak, Hibyan, & Webbe, 2014).

The most recent version of the SCAT, known as the SCAT3, evolved following the 4th International Conference on Concussion in Sport in 2012 (McCrory et al., 2013). Although only slight alterations were made, there was considerable evidence to support the revision (Guskiewicz et al., 2013). Importantly, the SCAT3 added medical background questions pertaining to potential factors that may affect baseline test interpretation or that may affect recovery. Such questions inquire about the athlete’s history of concussion, migraine headaches, diagnosed psychiatric disorder (or a family member diagnosed with a psychiatric illness), and a history of a diagnosed learning disability or attention deficit hyperactivity disorder (LD/ADHD). Although the SCAT3 is comprised of the same assessment measures as the SCAT2, the primary modifications included the removal of the SCAT2 total score and reordering of the assessment protocol. The SCAT2 total score was comprised of separate subcomponent scores taken from the subtraction of the total number of symptoms and errors on the m-BESS, and the addition of the scores from the physical signs score, the GCS, and the SAC, with a maximum score of 100 and with lower scores indicating poorer performance (Guskiewicz et al., 2013). The composite score therefore did not include the Maddocks score or the symptom severity score. As the sensitivity and utility of the SCAT2 total score was questioned, it was removed in the revised SCAT3,
which emphasized that clinicians interpret the results of each component of the entire protocol discretely (Guskiewicz et al., 2013).

2.6 Test validity

In consideration of the utility and sensitivity of any measure, it is important to introduce the concept of test validity, especially in its relation to SRC. Test validity refers to the ability of a test to accurately measure what it is intended to measure (Parikh, Mathai, Parikh, Sekhar, & Thomas, 2008). In the context of head injury, an assessment tool should be sensitive, in that it has a high probability of accurately identifying a concussed athlete as having sustained a concussion. The tool should also be highly specific, in that it has a high likelihood of identifying a healthy athlete as not having experienced a concussion (Putukian et al., 2014). The accuracy of such assessment tools are commonly referenced to a “gold standard” that establishes the true status of the injury (Parikh et al., 2008). As there is no specific gold standard for the diagnosis of concussion, the gold standard typically used to determine sensitivity and specificity in the sport context is the clinical diagnosis of concussion, which is grounded in a set of recommended guidelines, such as the American Academy of Neurology nomenclature (McCrea et al., 2003; 2005; Schatz & Sandel, 2013).

2.7 Modifying factors to concussion testing

Ensuring test validity is not the only property that must be considered when diagnosing concussion, with recent research outlining several potential population-specific factors that may obscure test interpretation (Dessy, Rasouli, Gometz, & Choudhri, 2014). Such moderating factors include age, gender, history of a learning disability or attention disorder, history of a psychiatric disorder, and a history of previous concussions (Dessy et al., 2014).

2.7.1 Age

Age has been found to significantly moderate baseline neurocognitive and balance evaluations, with younger athletes performing significantly worse in comparison to older athletes in an adolescent sample (Snyder & Bauer, 2014; Valovich McLeod, Bay, Lam, & Chhabra, 2012). Younger athletes have also been shown to endorse significantly more symptoms at baseline compared to older athletes (Valovich McLeod et al., 2012).
2.7.2 Sex
Sex should also be considered when interpreting test results, with research revealing females perform significantly better on the SAC and BESS at baseline, compared to males in an adolescent sample (Valovich McLeod et al., 2012). Reports on the effect of gender on symptom endorsement have varied, with some studies finding that females report significantly more symptoms and of greater severity at baseline compared to males (Covassin et al., 2006; Snyder & Bauer, 2014), and other studies asserting no differences in symptom endorsement between sexes (Jinguji et al., 2012; Valovich McLeod et al., 2012; Zimmer et al., 2014).

2.7.3 Learning disability, mental health, history of concussion
A history of a learning disability (LD) or attention deficit hyperactivity disorder (ADHD) is similarly believed to obscure test interpretation, with research revealing that such athletes perform significantly worse on baseline paper and pencil NP tests (Collins, Grindel, & Lovell, 1999). A history of a psychiatric disorder or the endorsement of psychiatric personality traits, such as depression or anxiety, also alters baseline concussion test analyses, with research demonstrating significantly greater symptom endorsement in such individuals (Bailey, Samples, Broshek, Freeman, & Barth, 2010; Balasundaram et al., 2016; Covassin, Elbin, Larson, & Kontos, 2012; Harmon et al., 2013). Finally, a history of concussion has shown to be associated with significantly greater symptom reporting at baseline (Register-Mihalik, Mihalik, & Guskiewicz, 2009; Valovich McLeod et al., 2012). These findings underscore the importance of baseline testing for accurate post-injury test interpretation, however if baseline data is lacking, such factors should be taken into consideration to guide appropriate clinical management.

2.8 Clinical management of concussion
Clinical management of concussion remains controversial, as no evidence-based guidelines exist (Harmon et al., 2013). Current protocols advocate for cognitive and physical rest during the acute post-injury period in order to prevent symptom exacerbation and delayed recovery (Harmon et al., 2013; Leddy et al., 2012; McCrory et al., 2013). However, there is limited research investigating what constitutes as ‘rest’, the optimal time of rest, as well as the long term outcomes of rest (McCrory et al., 2013). Indeed, prolonged rest has been suggested to have the potential of causing subsequent symptoms of fatigue, depression and physiological
deconditioning (Leddy et al., 2010). After symptoms have resolved at rest, athletes are encouraged to progress through a graduated return-to-play protocol that commences with light aerobic exercise and advances to sport specific activities, until the complete return to sport participation (McCrory et al., 2009; Willer & Leddy, 2006). If symptoms are reported following physical activity, it is recommended that further progress be discontinued until preceding activities are completed without symptoms (Harmon et al., 2013). Current guidelines endorse that the concussed athlete only return-to-play once asymptomatic and are able to progress through the exercise protocol without an exacerbation of symptoms (Leddy et al., 2010).

More recent research has targeted the role of exercise in managing prolonged recovery from concussion, with research demonstrating improved symptoms following controlled aerobic activity (Leddy et al., 2010; 2012). However, there is limited evidence at this time to endorse a specific exercise protocol as a treatment for persistent symptoms.

2.9 Rationale for current research

Presently, there are no stand-alone, objective measures available to identify and diagnose concussion. It is evident from the literature that early diagnosis is critical for informing clinical management and for reducing the potential risk of subsequent injury if the athlete is returned to play prematurely.

The SCAT3 is now widely acknowledged as the recommended tool used to evaluate sport-related concussion on the sideline and is now also widely used within concussion clinics across the recovery trajectory (McCrory et al., 2013). Although there is expert consensus and research to support the use of the separate SCAT3 subcomponents, there is limited evidence validating the clinical utility of the measure in its entirety over time. Recent research conducted by Putukian and colleagues (2015) investigated the utility of the SCAT2 at the acute phase of injury, finding that a 3.5-drop in the total score from baseline was 96% sensitive and 81% specific in detecting concussion. As such, the current research sought to extend this prior work by assessing the clinical utility of the SCAT3 components individually and as a global measure across time. Moreover, there is limited research validating the utility of this measure when baseline data are not available. Establishing the sensitivity and specificity of the SCAT3 under each circumstance
can provide meaningful information to clinicians that can aid in decision-making following concussion. Finally, there is no research currently available comparing the utility of the SCAT3 to other concussion assessment tools. By comparing the utility of the SCAT3 to a computerized NP concussion assessment battery, clinicians will gain a better understanding of when and to what magnitude each tool is most sensitive to detecting concussion deficits.

As research has also demonstrated population-specific factors that may obscure post-injury interpretation of baseline SCAT2 scores (Jinguji et al., 2012; Snyder & Bauer, 2014; Valovich McLeod et al., 2012; Zimmer et al., 2014) it is pivotal to update this information using the revised SCAT3. Another related issue with the current knowledge pertaining to the SCAT3 is that the majority of normative baseline data comprise scores from athletes in the United States. As there is a lack of research comparing baseline performance on the SCAT3 between American and Canadian athletes, it cannot be assumed that performance would be equivalent. Therefore, it is critical that a normative dataset for the SCAT3 is generated using a Canadian population. Together, these gaps in knowledge underscore the importance of developing both normative baseline SCAT3 values and establishing potential modifying factors in a Canadian population, as well as determining the clinical utility of the SCAT3 across time, in comparison to other assessment tools, and when baseline data is lacking.

2.10 Research objectives

Considering the rationale, the objective for the current study is two-fold:

1) First, the current research aims to evaluate the clinical utility of the SCAT3 by:
   a) Defining the sensitivity and specificity of the measures’ subtests independently and in combination across time;
   b) Comparing the sensitivity and specificity of the SCAT3 to a computerized neurocognitive concussion assessment tool;
   c) Investigating differences in the sensitivity and specificity of the SCAT3 in the presence and absence of baseline data;

2) Second, the research aims to generate a normative baseline database for the SCAT3 in a Canadian population, and to identify potential population-specific moderating factors on SCAT3 baseline values that may affect post-concussion test interpretation
Chapter 3
Normative SCAT3 Values and Modifying Factors in Varsity Athletes Playing a High-Risk Sport

3.1 Introduction

Normative data are of paramount importance for objective test interpretation, especially in the absence of baseline measures. Normative data helps characterize typical test performance in a distinct population that may be used as a reference to guide individualized clinical assessment. Data norms are commonly standardized across individual demographics such as age, sex, education, and nationality. Research calibrating data for tests used within the SCAT3 has revealed important factors that may influence baseline performance.

Specifically, studies have shown that on the Standardized Assessment of Concussion (SAC), females typically perform better than males in both adolescent (Jinguji et al., 2012; Valovich McLeod, Barr, McCrea, & Guskiewicz, 2006; Valovich McLeod, Bay, Lam, & Chhabra, 2012) and collegiate samples (Zimmer, Marcinak, Hibyan, & Webbe, 2014), and that in an adolescent sample, younger athletes perform worse than older athletes (Snyder & Bauer, 2014; Valovich McLeod, Bay, Lam, & Chhabra, 2012). A recent study has also revealed that in a varsity athletic sample, history of concussion did not significantly alter baseline SAC scores (Zimmer, Marcinak, Hibyan, & Webbe, 2014). Additionally, although research demonstrates that a history of a diagnosed learning disability (LD) or attention deficit hyperactivity disorder (ADHD) has been shown to negatively impact baseline neurocognitive performance (Collins et al., 1999), this has yet to be investigated using the SAC specifically.

Moreover, research examining baseline symptom reporting has highlighted moderating factors that should be considered when interpreting post concussion scores. The influence of sex has varied, with some studies reporting that females report more symptoms of greater severity at baseline compared to males (Covassin et al., 2006; Snyder & Bauer, 2014), with others revealing no differences on symptom reporting between sexes (Jinguji
et al., 2012; Valovich McLeod et al., 2012; Zimmer et al., 2014). In regards to age, it has been found that younger athletes typically endorse more symptoms at baseline compared to older athletes (Snyder et al., 2014; Valovich McLeod, Bay, Lam, & Chhabra, 2012). History of concussion has also demonstrated differential effects on baseline symptom reporting, with one study revealing that adolescents with a history of concussion report significantly more symptoms than those without (Valovich McLeod et al., 2012), however in another study conducted in a collegiate athletic sample no significant differences in symptom reporting between athletes with or without a history of concussion was found (Zimmer et al., 2014). Research has further demonstrated that a previous diagnosis of depression or depressive symptoms is linked to greater symptom endorsement prior to injury (Balasundaram et al., 2016; Putukian et al., 2014).

Lastly, research establishing normative data for the Balance Error Scoring System (BESS) in adolescents has revealed that males typically perform worse compared to females (Jinguji et al., 2012; Valovich Mcleod et al., 2012) and that older athletes make less balance errors on the BESS compared to younger athletes (Snyder & Bauer, 2014; T. Valovich McLeod et al., 2006; 2012). However in an additional study using a collegiate sample no significant differences between sex, age, or history of concussion on the BESS were found (Zimmer et al., 2014).

Such studies addressing population-specific factors influencing baseline SCAT3 component scores have provided clinicians with the opportunity to objectively interpret post-concussion results. However, current research is limited, with it being largely confined to athletes in the United States, with limited research investigating a Canadian population. As standardized norms do not yet exist for Canadians, researchers and clinicians are left to make comparisons between athletes in Canada to those from other countries. At the present time, there is no research that has sought to compare baseline concussion test performance between American and Canadian varsity athletic populations. Hence, the equivalence of performance on such measures cannot be assumed, warranting the need to establish Canadian normative data.
3.2 Rationale
As baseline testing may not always be available to athletes, or similarly, as baseline scores may be invalid due to poor effort, fatigue, or illness at the time of testing (Shuttleworth-Edwards, Whitefield-Alexander, Radloff, Taylor, & Lovell, 2009), it is imperative that a representative database for SCAT3 scores is established. The limited normative data for the SCAT3 places the validity of this tool into question until further research can establish such standards (McCrory et al., 2013). It is therefore critical that a normative dataset for the SCAT3 is generated, which considers previous research identifying factors that may affect baseline concussion evaluation measures. Similarly, it is imperative that SCAT3 test scores are standardized in a Canadian athletic sample, given the limited research on this measure in Canada.

3.3 Research objective
Given the gaps in knowledge, the objective for the study was to define a normative database for SCAT3 component scores and to determine any modifying factors to the SCAT3 at baseline that could affect post-concussion test interpretation.

3.4 Methods and procedure
3.4.1 Design
The study was conducted using a cross-sectional research design, such that athletes were assessed on a number of different measures at one time-point.

3.4.2 Participants
A total of 524 varsity athletes participating in high-risk sports at the University of Toronto between August of 2014 and December of 2015 participated in this study. Participants were recruited in person during mandatory baseline testing from the University of Toronto’s varsity athletics teams, including men and women’s ice hockey, lacrosse, volleyball, soccer, rugby, and basketball, men’s football, baseball, and wrestling, and women’s field hockey. Athletes were eligible to participate in the study if
they were able to speak and understand English fluently, if they provided informed consent, and if they had no history of neurological or balance disorders. Prior to participation in this study, all athletes provided informed consent as approved by the Health Sciences Research Ethics Board (REB) at the University of Toronto. The information sheet and consent documents taken from the REB application can be found in Appendix 1.

3.4.3 Outcome measures

The main outcome measures were comprised of the major SCAT3 subtests. These tests included: (1) the Post Concussion Symptom Scale, (2) the SAC, and (3), the m-BESS. A copy of the SCAT3 can be found in Appendix 2.

1. **Post Concussion Symptom Scale**

The Post Concussion Symptom Scale (PCSS) is a self-report graded symptom checklist used to detect concussion (Guskiewicz et al., 2013). Although there are several variations of the PCSS, the questionnaire included in the SCAT3 contains 22 items on a 7-point Likert rating scale, ranging from 0 (asymptomatic) to 6 (severely symptomatic) (Okonkwo, Tempel, & Maroon, 2014). Each item characterizes a common sign or symptom related to concussion, such as dizziness, headache, irritability, and fatigue (SCAT3, 2013). The PCSS results in two scores: 1) the total symptom score that comprises of the total number of symptoms reported with a total possible score of 22, and 2) the symptom severity score that comprises of the sum of symptom severity across the 22 symptoms, for a possible maximum score of 132 (SCAT3, 2013).

2. **Standardized Assessment of Concussion**

The Standardized Assessment of Concussion (SAC) is an instrument designed to assess neurological status and neurocognitive impairment on the sideline and evaluates three cognitive domains: orientation, concentration, and memory. The orientation domain examines whether the athlete can recall the date and time of day, resulting in a maximum score of 5 (SCAT3, 2013). The concentration component requires the athlete to recite a
series of digits backwards, as well as months of the year in reverse order, with a range in score from 0 to 5, with 5 indicating better performance (SCAT3, 2013). Immediate and delayed memory is assessed by the athlete recalling a series of words immediately (score of 0-15) and after a short delay (score of 0-5), respectively (SCAT3, 2013). The total score for the SAC therefore ranges from 0-30, with 30 representing the maximum score that can be attained. The total test requires 5 to 7 minutes to administer and was designed for use by clinicians or athletic trainers with no previous experience in psychometric testing (Valovich McCleod et al., 2006).

3. Modified Balance Error Scoring System (m-BESS)

The Balance Error Scoring System (BESS) is an objective measure designed for use on the sideline to detect balance deficits characteristic of concussion (Onate, Beck, & Van Lunen, 2007). The original BESS is comprised of three 20-second duration balance assessments on both a foam and hard test surface. However, the modified version (m-BESS) used in the SCAT3 only evaluates the athlete’s balance on a firm test surface. The three poses assessed include a double leg stance, where the athlete stands with his/her feet together, hands on the hips, and eyes closed; a single leg stance, where the athlete stands on his/her non-dominant foot, with the dominant foot being held with approximately 45 degree knee flexion; and a tandem stance, where the athlete stands with the non-dominant foot directly behind the dominant foot (SCAT3, 2013). The m-BESS is scored across the three stances by subtracting the number of errors made by the athlete from the maximum possible amount of errors for each stance, 10, with the total score being calculated by subtracting the number of errors made from the total possible amount of errors on the test, 30. Therefore, a higher score indicates greater postural performance. Errors include lifting hands off of the iliac crest, opening eyes, step, stumbling, or falling, moving hip into greater than 30 degree abduction, lifting forefoot or heel, or remaining out of the test position for greater than 5 seconds (SCAT3, 2013).

3.4.4 Procedure and data collection

All baseline data were collected from new or returning players who had not previously completed baseline testing across the 2014-2015 and 2015-2016 athletic seasons. All
testing was completed in the concussion laboratory at the University of Toronto. Athletes were typically tested in one session with a group of teammates, rotating amongst three to four research assistants or graduate students trained in administering the SCAT3 and a range of vision assessments part of a larger scale study. If an athlete could not come into the lab with his/her other teammates due to scheduling difficulties or other unforeseen circumstances, the athlete was tested individually at another time close in proximity. The evaluations took approximately one hour to complete, with the SCAT3 lasting around 10 minutes.

3.4.5 Data analysis

All analyses were performed using IBM Statistical Package for the Social Sciences Statistics (SPSS) Version 23. Descriptive statistics were performed to determine the frequency and central tendency of the major SCAT3 scores across the variables of sport, sex, years of education, history of a learning disability or attention deficit disorder, history of depression, anxiety or other psychiatric disorder, a family history of a psychiatric disorder, history of migraine headaches, and history of concussions.

Tests of normality were conducted by evaluating skewness and kurtosis, as well as through visualization of Q-Q plots. As data from the SCAT3 subcomponents were non-normally distributed, the appropriate nonparametric data analyses were performed.

Specifically, to investigate the effect of each binary independent variable, including sex (male/female), history of migraines (yes/no), history of mental health (yes/no), family history of mental health (yes/no), and history of a learning disability/ADHD (yes/no) on SCAT3 subcomponent scores separate Mann-Whitney U tests were performed. The Mann-Whitney U test is an alternative to the independent samples t-test for nonparametric data that compares the medians of a dependent continuous variable between two groups. If the variance of the test scores is similar across groups, the medians are compared for significance. However, if the variance is different across groups, the median rank is evaluated for significance. Therefore, although it is not intuitive, a significant difference in test scores may be found between groups even if the medians of
the groups are the same or are very similar. Levene’s test of homogeneity of variance was used to determine whether distribution of scores were equal across groups for each independent variable. Effect sizes were computed as $\eta^2$ in order to describe the amount of variance accounted for in the dependent variable by the independent variable.

To investigate the effect of years of education and history of concussion on the SCAT3 scores, separate Kruskal Wallis nonparametric tests were performed with effect sizes calculated as $\eta^2$. The Kruskal Wallis test is another rank-based nonparametric test that is analogous to a One-Way analysis of variance (ANOVA) that can be used to determine whether there is a statistically significant difference between two or more groups on a dependent variable. This test similarly evaluates significance of the medians if the variance between groups is similar, or of mean ranks if the variance is different across groups. If a significant main effect was found on the Kruskal Wallis test, post hoc pairwise analyses were performed with Dunn-Bonferonni adjustments.

As there is no known analysis of covariance (ANCOVA) for data with a non-normal distribution, the independent variables that were found significant in the preliminary analyses were stratified with the remaining independent variables in order to determine whether significant results were driven by such covariates. In order to control for the risk of committing a multiple-comparisons error, covariate analyses were only conducted if significant differences emerged in the preliminary analyses of the SCAT3 component scores. However, out of interest, the interaction of concussion history was assessed across all other significant variables. The stratified variables were entered into a Kruskal Wallis test, with follow-up pairwise analyses being performed with Dunn-Bonferonni corrections to determine where exact differences occurred while accounting for unequal sample sizes. Subsequent Chi square tests were computed to determine whether any associations between the independent variables arose. Additionally, effect sizes were computed as $\eta^2$ in order to describe the amount of variance accounted for in the dependent variable by the independent variable, while controlling for the other variable, with 0.01 indicating a small effect size, .06 a medium effect size, and .14 as a large effect.
3.5 Results

3.5.1 Demographics

At the time of testing, athletes were between the ages of 17 and 24 (M= 19.8, SD= 1.93), females made up 36.3% of the sample, 36.4% athletes reported a history of 1-2 previous concussions, with 10.2% of the sample reporting 3 or more previous concussions, 7.3% of the sample reported a history of migraine headaches, 4.8% of the sample had a previous diagnosis of a learning disability or ADD/ADHD (LD/ADHD), and 8.3% and 18.3% of the sample reported having personally been diagnosed or having had a family diagnosed with depression, anxiety, or another psychiatric disorder, respectively. Football had the highest frequency of athletes completing testing, followed by rugby. Table 3.1 provides a description of the frequency and percentage of participants across all sports. The majority of participants completed either 12 years of education (33.6%) or 15 years of education or greater (30.5%). A complete description of the frequency and percentage of participants by year of education is presented in Table 3.2.
Table 3.1. Number of participants per sport

<table>
<thead>
<tr>
<th>Sport</th>
<th>N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Football</td>
<td>121 (23.1)</td>
</tr>
<tr>
<td>Rugby</td>
<td>84 (16)</td>
</tr>
<tr>
<td>Ice Hockey</td>
<td>66 (12.6)</td>
</tr>
<tr>
<td>Soccer</td>
<td>63 (12)</td>
</tr>
<tr>
<td>Lacrosse</td>
<td>56 (10.7)</td>
</tr>
<tr>
<td>Basketball</td>
<td>48 (9.2)</td>
</tr>
<tr>
<td>Volleyball</td>
<td>42 (8)</td>
</tr>
<tr>
<td>Field Hockey</td>
<td>24 (4.6)</td>
</tr>
<tr>
<td>Baseball</td>
<td>13 (2.5)</td>
</tr>
<tr>
<td>Wrestling</td>
<td>7 (1.3)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>524</strong></td>
</tr>
</tbody>
</table>

Table 3.2. Number of participants per year of education

<table>
<thead>
<tr>
<th>Years of Education</th>
<th>N(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>174 (33.6)</td>
</tr>
<tr>
<td>13</td>
<td>107 (20.7)</td>
</tr>
<tr>
<td>14</td>
<td>79 (15.3)</td>
</tr>
<tr>
<td>15+</td>
<td>158 (30.5)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>518</strong></td>
</tr>
</tbody>
</table>

3.5.2 Modifying factors

I. Sex

A Mann-Whitney U test was run to determine if there were significant differences in total symptom score, symptom severity score, SAC score, and m-BESS score between males and females. Distributions of total symptom and symptom severity scores were similar for males and females, however the variances of the SAC score were significantly different between sexes, as evaluated by Levene’s statistic (p < .05). Sex significantly moderated the baseline total symptom
score, with males reported more symptoms (mdn= 2, mean rank= 274.86) compared to females (mdn= 1.5, mean rank= 240.77), $U = 27601$, $z = -2.517$, $p = .012$, $\eta^2 = .007$ (Figure 3.1). This finding remained significant for the symptom severity score, with males reporting symptoms of greater severity (mdn= 3, mean rank= 273.56) compared to females (mdn= 2, mean rank= 243.06), $U = 28035.5$, $z = -2.517$, $p = .025$, $\eta^2 = .004$ (Figure 3.2).

Figure 3.1. Median PCSS total symptom score greater in males compared to females.

Figure 3.2. Median PCSS symptom severity score greater in males compared to females.
The SAC score was also found to be significantly different between males and females, with females performing better on the SAC (median=27, mean rank=284.29), compared to males (median=27, mean rank=250.1), $U = 35871$, $z = 2.513$, $p = .012$, $\eta^2 = .019$. This data is presented graphically in Figure 3.3. Lastly, there were no significant differences on the balance score between males and females.

![SAC Total ranked greater in females](image)

*Figure 3.3. Mean ranking of SAC performance greater in females compared to males.*

II. History of migraines

There were no significant differences between individuals with a history of migraines compared to athletes without, across any of the SCAT3 subtests as revealed by Mann-Whitney U test analyses.

III. Previous diagnosis of depression, anxiety, or other psychiatric disorder

The variances in total symptom scores and symptom severity scores were significantly different across athletes with a history of mental health and those without, as evaluated using Levene’s test of equality ($p < .05$); therefore mean ranks were compared for significance. The Mann-Whitney U test revealed significantly greater total symptoms in athletes with a history of mental health (mdn=5, mean rank=316.94), compared to those without (mdn= 2, mean rank= 255.41),
U= 12682.5, z= 2.613, p= .009, η²= .023 (Figure 3.4). Similarly, athletes with a history of mental health (mdn=6, mean rank= 318.12) reported symptoms of greater severity compared to those without (mdn= 3, mean rank= 255.31), U= 12733, z= 2.663, p= .008, η²= .032 (Figure 3.5).

There were no significant differences between groups with and without a history of mental health on SAC scores or balance measures.

**Figure 3.4.** Mean ranking of PCSS total symptoms greater in athletes with a history of mental health compared to those without.
IV. Family history of mental health

A significant positive correlation was found between athletes with a history of mental health and athletes with a family history of mental health, $\phi = 0.440, p < .001$. However, for completeness, a Mann-Whitney U test was performed to assess its effect on SCAT3 scores. As distributions were dissimilar across groups for the total symptom score and symptom severity score ($p < .05$), mean ranks were evaluated for significance. As predicted, athletes with a family member previously diagnosed with a mental health disorder reported significantly more symptoms (mdn= 3, mean rank= 304.3) of greater severity (mdn= 4.5, mean rank= 306.16), compared to those without (TS: mdn= 2, mean rank= 250.84; SS: mdn= 2, mean rank= 250.42), $U= 24139, z= 3.172, p= .002, \eta^2= .02$ (Figure 3.6), $U= 24314, z= 3.302, p= .001, \eta^2= .027$ (Figure 3.7), respectively. Similarly, there were no significant differences found between groups on the SAC or m-BESS score.
Figure 3.6. Mean ranking of PCSS total symptoms greater in athletes with a familial mental health history compared to those without.

Figure 3.7. Mean ranking of PCSS total symptoms greater in athletes with a family history of mental health compared to those without.
V. Previous diagnosis of a learning disability or ADD/ADHD

Distributions of scores amongst athletes with and without a history of a learning disability (LD) or ADD/ADHD were similar across all tests, as demonstrated by Levene’s test of homogeneity (p > .05). Total symptom score, symptom severity score, and balance score were all non-significant between groups. However, the median SAC score was significantly lower in athletes with a history of LD or ADHD (mdn= 26), compared to athletes without a previous diagnosis (mdn= 27), U = 4523, z=-2.296, p= .022, $\eta^2$=.005 (figure 3.8).

![Figure 3.8. Median SAC score lower in athletes with a history of a learning disability or ADHD compared to athletes without.](image)

VI. Years of Education

Age and years of education were significantly correlated, r=.797, p < .001; therefore years of education was analyzed exclusively in supplementary investigations. Levene’s statistic revealed significantly different distributions across groups for the total symptom score, the symptom severity score, and the balance score (p < .05). As such, mean ranks were evaluated for significance for these tests. Results from the Kruskal Wallis analysis demonstrated a significant difference between groups on the total symptom score, $\chi^2(3)$= 7.996, p= .046, $\eta^2$=.028. Post hoc
analyses revealed no significant differences across years of education on total symptom scores, however a trend was revealed such that athletes with 12 years of education reported higher symptoms (mdn = 3, mean rank = 284.58) compared to athletes with 15 or more years of education (mdn = 2, mean rank = 243.74), p = .069. No other significant differences were found within the SCAT3 subtests amongst the different years of education.

VII. Concussion History
Concussion history was binned into 3 categories: 0, 1-2, or 3 or more. The distribution of symptom scores and m-BESS scores were similar as assessed by Levene’s statistic (p > .05), however the distribution of SAC scores was significantly different across groups (p < .05). No significant differences were found within any of the SCAT3 subtests between athletes with and without a history of concussion.

3.5.3 Covariates

I. Variables Interacting with Sex

Sex and History of Learning Disability or ADHD
As there were unequal variances between males and females with and without a history of a LD/ADHD on the SAC score (p < .05), mean ranks were compared for significance. All other tests demonstrated equal variances. Differences on the SAC score were significant, $\chi^2 (3) = 11.866$, p = .008, with post hoc tests revealing that males with a history of a diagnosed LD/ADHD (mean rank = 171.62) performed significantly worse than females without a previous diagnosis (mean rank = 284.2), p = .017. The size of the interaction effect was small, $\eta^2 < .0001$, with sex accounting for 0.4% of the variance and history of LD/ADHD accounting for 0.29% of the variance in SAC scores.

Sex and History of Mental Health
Distributions of scores on the SAC and symptom measures were significantly different between males and females with and without a history of mental health (p < .05). The scores on the m-BESS were equally distributed across groups. A significant main effect was found for the SAC, $\chi^2 (3) = 9.010$, p = .029, total symptoms, $\chi^2 (3) = 10.686$, p = .014, and symptom severity scores, $\chi^2 (3) = 9.892$, p = .02. Follow-up analyses revealed that total symptoms and symptoms severity
scores were significantly greater in males with a history of mental health (mean rank: TS=334.52; SS=333.3), compared to females without (mean rank: TS: 236.98, p = .022; SS: 238.65, p = .029). Mental health history accounted for more of the variance in total symptoms score (2.4%) and symptom severity score (3.2%), compared to sex (TS: .7%, SS: .4%).

Post hoc analyses of the SAC main effect revealed no significant differences between groups, however differences in SAC scores approached significance between males with a history of mental health (mean rank= 244.55), compared to females without (281.24), p= .059. A history of mental health also accounted for more of the variance in the SAC compared to sex, however this effect was small (.27%).

**Sex and History of Concussion**
Test scores were equally distributed for the m-BESS, total symptoms and symptom severity; however were unequal for the SAC (p < .05). There were no significant differences across males and females with 0, 1-2, or 3 or more concussions on the m-BESS or symptom severity score, however symptom severity approached significance, $\chi^2 (5)= 10.806$, p = .055. Differences in total symptom scores was found to be significant, $\chi^2 (5)= 12.872$, p = .025, with follow-up analyses revealing that males with 1-2 previous concussions (mean rank= 287.37) reported significantly more symptoms at baseline compared to females with no history of concussions (mean rank= 222.82), p = .013. Concussion history accounted for more of the variance in total symptom score compared to sex, however this effect was small (.58%).

A significant main effect was also found for the SAC, $\chi^2 (5)= 12.857$, p = .025, with post hoc tests revealing that males with no history of a concussion (243.33) performed significantly worse than females with 1-2 previous concussions (mean rank= 317.83), p = .009. The effect size was greatest for sex, accounting for 1.5% of the variance in the SAC score, compared to concussion history (.44%).

**Sex and Years of Education**
Levene’s test of homogeneity revealed that all tests were unequally distributed (p < .05); therefore mean ranks were compared for significance. A significant effect was found between males and females across years of education for total symptom score, $\chi^2 (7)= 16.922$, p = .018,
and symptom severity score, \( \chi^2 (7) = 14.925, p = .037 \), with post hoc tests revealing no specific differences between groups. Years of education accounted for most of the variance in the total symptom score (2.3%) and symptom severity score (1.5%), compared to sex (TS: .97%, SS: .69%). Neither SAC nor m-BESS scores were significantly different across groups.

II. Variables Interacting with History of Mental Health

**History of Psychological Illness and Learning Disability**

Test scores were unequally distributed across athletes with and without a history of mental health and/or a learning disability for the total symptom and symptom severity measures (p < .05). The variances of scores were equal across groups for the SAC and m-BESS, and therefore medians were compared for significance. A significant main effect was found for the SAC, \( \chi^2 (3) = 12.781, p = .005 \), with post hoc analyses revealing significantly lower SAC scores for athletes with a history of a learning disability only (mdn= 26) in comparison to athletes with no history of either LD/ADHD or mental health (mdn= 27), p = .028, to athletes with a mental health history only (mdn= 27), p = .045, and compared to athletes with a history of both LD/ADHD and mental health (mdn= 29.5), p = .029. The size of this effect was greatest for LD/ADHD, accounting for 1.5% of the variance compared to a history of a mental health disorder (.13%). There were no other significant differences found across groups for the symptom or balance measures.

**History of Psychological Illness by Number of Previous Concussions**

Distributions were equal for the total symptom score, m-BESS, and SAC, however Levene’s test of homogeneity revealed significantly unequal variances for the symptom severity scores across groups (p < .05).

A significant main effect was found for the total symptom score, \( \chi^2 (5) = 15.214, p = .009 \), and the symptom severity score, \( \chi^2 (5) = 14.929, p = .011 \). Post hoc analyses revealed that total symptoms and symptom severity scores were significantly greater in athletes with a history of a mental health disorder and 1-2 previous concussions (TS: mdn= 4; SS: mdn= 6, mean rank= 335.3), compared to athletes with no previous concussions or a diagnosis of a psychiatric disorder (TS: mdn= 2, p = .009; SS: mdn= 2, mean rank= 240.28, p = .006). A history of a psychological illness accounted for more of the variance in the total symptom score (.97%) and symptom severity
score (1.4%), compared to concussion history (TS: .60%, SS: .98%). No other significant differences were found between groups on the SAC or m-BESS.

A significant positive correlation was also revealed between athletes with a history of mental health and number of previous concussions, $\phi = .112, p= .04$, suggesting that as the number of previous concussions increases, the risk of mental health problems also increases, however a causal interaction cannot be concluded.

**History of Psychological Illness and Years of Education**

Distributions of scores were significantly different across groups for the symptom measures and the SAC, as assessed by Levene’s test of homogeneity ($p < .05$). A significant main effect was found for total symptom score, $\chi^2 (7)= 18.765, p= .009$ and symptom severity score, $\chi^2 (7)= 18.013, p= .012$ with post hoc analyses revealing significantly greater symptom reports in athletes with a history of a psychological illness with 12 years of education (TS: mean rank= 405; SS: mean rank= 406.58), compared to athletes without a mental health history with 13 (TS: mean rank= 243.95, p= .009; SS: mean rank= 243.6, p = .008), 14 (TS: mean rank= 251.02, p= .023; SS: mean rank: 254.94, p= .028), and 15 or more years of education (TS: mean rank= 238.98, p= .005; SS: mean rank= 239.63, p= .005). Years of education accounted for more of the variance in both total symptom score (3.3%) and symptom severity score (2.4%), compared to a history of a diagnosed psychological illness (TS: 1.9%, SS: 2.3%). No other significant differences were found across groups for the SAC or m-BESS.

### III. Learning Disability Interactions

**History of a Learning Disability and Number of Previous Concussions**

There were no significant differences found on any of the subtests for athletes with a previous diagnosis of a learning disability and previous concussions, however, SAC scores approached significance, $\chi^2 (5)= 10.973, p=.052$, with LD/ADHD accounting for most of the variance (.36%).

**History of a Learning Disability and Years of Education**

Distributions of test scores were significantly different across groups for the symptom measures and the m-BESS, as revealed by Levene’s test of homogeneity ($p < .05$); therefore mean ranks
were compared for significance. Total symptom scores, $\chi^2 (7)= 14.617, p= .041$, and symptom severity scores, $\chi^2 (7)= 15.276, p= .033$, were significantly different between athletes with and without a learning disability across the years of education, however post hoc analyses found no significant differences between specific groups.

The effect size was largest for the interaction between variables, accounting for 2.3% of the variance in total symptoms, and 3.1% of the variance in symptom severity score. Independently, years of education accounted for more of the variance in total symptom (1.97%) and symptom severity (2.99%) scores than a history of LD/ADHD (TS: .24%; SS: .38%). No other significant differences were found between groups on the SAC or m-BESS.

IV. Interactions with Concussion History

Concussion History and Years of Education

The distribution of test scores was significantly different across groups for the symptom and balance measurement ($p < .05$), however were similar for the SAC ($p > .05$), as revealed by Levene’s statistic. There were no significant differences between groups on any of the SCAT3 subtests.

3.6 Discussion

The current study provides the first demographically normal data for the primary SCAT3 component scores in a Canadian varsity athletic sample across the factors of: sex, years of education/age, history of migraine headaches, history of a diagnosed psychiatric disorder personally or within the family, previous diagnosis with LD/ADHD, and history of concussion.

The results reported in this study are largely consistent with previous research; however also elucidate earlier findings by highlighting important interactions between population-specific factors influencing baseline SCAT3 scores.

Specifically, the role of sex in the literature has demonstrated mixed findings regarding its influence on symptom reporting in athletes during baseline testing (Covassin et al., 2006; Jinguji et al., 2012; Snyder & Bauer, 2014; Valovich McLeod et al., 2012; Zimmer et al., 2014). While
the current study revealed that symptom reporting was significantly greater in males compared to females, the interaction analyses demonstrated that years of education, history of a psychological illness, and concussion history explained more of the variance in this result than sex. In particular, male athletes with a history of a mental health disorder reported more symptoms of greater severity compared to females without a previous diagnosis, and similarly, males with 1-2 previous concussions reported more symptoms at baseline compared to females without a history of concussion. Although the interaction between years of education and sex on symptom scores was significant, follow-up analyses did not inform where such differences were. Comparison of effect sizes for the interactions between these three variables revealed that years of education, followed by mental health history accounted for most of the variance amongst total symptom and symptom severity scores compared to concussion history. When analyzed independently, however, mental health history accounted for most of the variance in symptom scores. This finding elucidates the inconsistency in the literature regarding the role of sex on concussion symptom reports, while underscoring the significance of an athlete’s past and current mental health, as well as education in augmenting baseline symptom scores.

The role of sex on cognition, however, appears more direct. The results of this study revealed that females performed significantly better than males on the SAC. While concussion history and sex significantly interacted, with males with 1-2 previous concussion performing significantly worse than females without any previous concussions, this effect was largely driven by sex. Additionally, although a history of LD/ADHD was associated with reduced SAC scores, a significant follow-up interaction between a history of LD/ADHD and sex revealed that more of the variance in the SAC scores was accounted for by sex. This finding is consistent with previous research demonstrating poorer performance in males on the SAC compared to females in a collegiate sample (Zimmer et al., 2014); however this study is the first of its kind to reveal the lesser role of concussion history and LD/ADHD in driving such results.

The effects of a family member previously diagnosed with a psychiatric disorder on PCSS scores were similar to those found for athletes with a personal mental health history. Specifically, athletes with a family member previously diagnosed with a psychiatric disorder reported more symptoms of greater severity compared to those without. Additionally, a significant positive correlation was discovered between an athlete’s personal diagnosis and a family member’s
diagnosis of a psychiatric illness. As psychiatric disorders have been linked to a genetic basis, it is reasonable that an association was found (Lee, Ripke, Neale, Faraone, Purcell, et al., 2013).

Another important association found amongst the interacting variables was the positive correlation between athletes with a history of concussion and a diagnosed psychiatric disorder, suggesting that as the number of concussions increases, as does the risk for a mental health disorder, however a causal explanation cannot be inferred. This finding is consistent with previous research demonstrating the association between depression risk and concussion history in adolescents, as well as in retired professional football players (Chrisman & Richardson, 2014; Guskiewicz et al., 2007). This result therefore provides further evidence to support the claim of the detrimental psychological impacts of concussion throughout the lifespan and underscores the importance of early detection using such sideline measures such as the SCAT3 to guide appropriate management.

In regards to the effect of a history of a psychiatric disorder on baseline SAC scores, the findings were mixed. Although previous research has suggested that the presence of a psychiatric disorder can be associated with reduced cognitive abilities, the preliminary analyses of this study assessing the individual effect of a previous mental health diagnosis on SAC scores found no significant differences between athletes with or without a previous diagnosis (Millan et al., 2012). This may be due to the brevity of the test with a lesser ability to detect subtle cognitive differences.

Follow-up interaction analyses, however, revealed a paradoxical finding between a previously diagnosed psychiatric disorder and LD/ADHD on the SAC, such that athletes with a history of only LD/ADHD performed significantly worse than athletes without either diagnoses, in comparison to athletes with a previous mental health diagnosis only, as well as in comparison to athletes with both a history of mental health and LD/ADHD. Although it is conceivable that an athlete with a history of LD/ADHD would perform worse than athletes without the presence of either diagnoses or than athletes with only a mental health history, it is unexpected that this group would perform significantly better than athletes with both a history of mental health and LD/ADHD. Although a Dunn-Bonferonni correction was applied for post-hoc analyses following the Kruskal-Wallis test, accounting for unequal sample sizes, the small number of
athletes with both a mental health history and LD/ADHD had surprisingly high SAC scores. This finding may not be generalizable to a greater population, and therefore future research investigating the combined effect of mental health history and LD/ADHD is warranted. Nevertheless, for the interaction, the variable of LD/ADHD accounted for more of the variance in the SAC scores compared to the mental health variable, suggesting that a diagnosis of LD/ADHD is a more important factor to consider while interpreting post-concussion cognitive data.

Lastly, in regards to the balance evaluation, although previous research has found sex and age differences on the m-BESS in an adolescent sample (Jinguji et al., 2012; Snyder & Bauer, 2014; Valovich McLeod et al., 2006; 2012), the current study found consistent results with research using a collegiate sample where neither sex, age, nor history of concussion significantly influenced baseline m-BESS scores (Zimmer et al., 2014). The current study further contributed to these findings demonstrating that neither a history of migraine headaches, LD/ADHD, or family history of a psychiatric disorder significantly moderated baseline balance functioning as assessed with the m-BESS.

Study Limitations
This study delineated important population-specific factors moderating major baseline SCAT3 component scores; however this study was not without its limitations.

Methodological issues
One potential limitation that must be considered is that the SCAT3 consisted of only a small portion of a breadth of baseline tests administered, with the length of testing potentially causing fatigue or mental strain that could have affected test scores. As previous research has demonstrated that fatigue increases symptom reporting and reduces balance performance during baseline testing, all efforts were made to ensure that the SCAT3 was administered early on in the testing battery to reduce the extent of this influence (Piland, Ferrara, Macciocchi, Broglio, & Gould, 2010; Wilkins, Valovich McLeod, Perrin, 2004).
Another potential limitation of this study was that the collection of baseline data utilized a rotational approach within a larger group of athletes that may have produced distraction when completing the tests. However, all athletes were instructed that the baseline evaluations should be considered analogous to a medical test and taken seriously in order to ensure that complete attention and effort was provided on all of the tests.

Statistical considerations
Another important consideration that must be addressed is the clinical interpretation of the findings given the statistical approach used. Specifically, when the variances between groups were unequal, the mean ranking of the medians were compared for significance rather than the exact median values. While investigating the effect of sex on SAC scores, it was found that males performed significantly worse compared to females using this statistical approach, however, the median values were the same. Although this may not appear clinically relevant, nonparametric tests are suggested to have less statistical power than parametric tests (Hollander, Wolfe, & Chicken, 1973), with this finding therefore confirming the sizeable impact of this effect in a clinical setting. Additionally, after analyzing this data using a parametric approach, the finding of females (M= 26.96) performing better than males (M= 26.27) on the SAC remained significant, t= -3.448, p= .001, further providing evidence for the clinical significance of this result.

The results should also be interpreted with caution due to the lack of controlling for multiple comparisons. Specifically, family-wise error rate is the inflation in the probability of making a Type I error, or incorrectly rejecting the null hypothesis, when performing multiple hypotheses tests. The current study investigated the effect of sex, years of education, history of concussion, family and personal history of mental health, history of a diagnosed learning disability, and history of migraine headaches on SCAT3 scores. Although a more conservative approach may be used to control for multiple comparisons by applying Bonferonni corrections, only post-hoc analyses after finding a significant main effect implemented this statistical approach. This decision was guided by current statistical recommendations against the use of a more conservative approach, as it is rarely applied consistently, causing problems for interpretation in meta-analyses (Hopkins, Marshall, Batterham, & Hanin, 2009). Indeed, many recent studies investigating normative baseline SCAT2/3 scores have not controlled for multiple comparisons.
In addition to this, non-parametric tests have less power than parametric approaches, therefore reducing the chance of finding a significant effect. As such, a less conservative approach was employed. Overall, although the results should be considered with caution, the key findings of the moderating factors on baseline SCAT3 scores replicated previous research so should continue to be considered salient.

**SCAT3-specific limitations**

The current research also reveals some of the limitations characteristic to the SCAT3. Specifically, the current study examined the combined effect of LD/ADHD on the major SCAT3 component scores- a variable that was taken directly from the background questions inquired on the SCAT3. Although LD and ADHD were grouped together as one variable in this study, recent research has revealed that a diagnosis of LD alone and in combination with a diagnosis of ADHD is associated with lower cognitive performance on a computerized testing battery in comparison to athletes only diagnosed with ADHD (Elbin et al., 2013). This finding underscores the importance of examining for the presence of LD and ADHD independently and should be considered when administering the SCAT3.

Furthermore, a history of concussion was not found to significantly moderate the major SCAT3 component scores when investigated independently from the other variables. However, it is difficult to attain an athlete’s true concussion history as it is often measured through self-report and is subject to recall bias. The SCAT3 also only inquires on the number of previously diagnosed concussions and given that an athlete might not always seek medical attention following a suspected concussion, the number of concussions reported might be underestimated. Additionally, the severity of previous concussions was not assessed as a potential factor influencing baseline test scores in this study. The background questions integrated in the SCAT3 inquire on how recent an athlete’s previous concussion was, as well as the length of recovery from the most recent concussion, providing insight into the severity of the most recent injury. Therefore, future research investigating the effect of previous concussion severity on baseline SCAT3 scores is warranted.
In a similar vein, this research identified that a history of mental health was associated with increased symptom reporting on baseline PCSS scores. One limitation to this study and to the SCAT3 specifically, is that no objective measure was used to assess mental health functioning. Instead, in the background questions of the SCAT3, it is asked, “Have you ever been diagnosed with depression, anxiety or other psychiatric disorder?” As such, the results from this research are unclear as to which specific mental health diagnosis leads to increased symptom reporting. The addition of a more objective test, such as the Beck anxiety (Beck, Ward, Mendelson, Mock & Erbaugh, 1961) or depression inventory (Beck, Ward, Mendelson, 1961) at baseline should be considered to elucidate these findings and contribute to post-concussion test interpretation.

3.7 Conclusion

The results from this study contribute meaningful information to the field of concussion research and clinical practice.

Firstly, this research established a large normative dataset for the major SCAT3 components in a Canadian varsity athletic population. Currently, norms are available for the SCAT3 components in an American population (Jinguji et al., 2012; Snyder & Bauer, 2014; Valovich McLeod, Barr, McCrea, & Guskiewicz, 2006; Valovich McLeod, Bay, Lam, & Chhabra, 2012, Zimmer, Marcinak, Hibyan, & Webbe, 2014). However, there is no previous research evaluating differences in SCAT3 scores between Canada and the United States. Variability in scores may arise due to inherent distinctions between American and Canadian sport culture, especially in regards to the increased academic merit at the collegiate level in Canada. As such, baseline SCAT3 norms established in the United States may not be appropriate for comparison against athlete’s post-concussion scores in Canada. The normative database of SCAT3 scores established from a Canadian population in this study is therefore highly meaningful.

Importantly, this study also highlights population-specific factors moderating baseline SCAT3 scores that may affect post-concussion test interpretation.

It was found that sex explained most of the variance in SAC scores, with males performing more poorly than females, followed by a previous diagnosis of LD/ADHD, with athletes with a history
of LD/ADHD performing worse than athletes without. Mental health also accounted for the most variance in total symptom and symptom severity scores of the PCSS, with athletes previously diagnosed with a psychiatric disorder reporting more symptoms of greater severity at baseline compared to athletes without previous mental health issues. Having a family member previously diagnosed with a psychiatric disorder also increased baseline symptom reporting, however the positive correlation found with athletes personally diagnosed with a psychiatric illness may have driven such results. Years of education also arose as a potential moderator of baseline PCSS scores, with athletes with 12 years of education reporting more symptoms of greater severity compared to athletes with more years of education. As recent research has revealed an association between heightened baseline PCSS scores and worse outcomes following concussion (Custer et al., 2016), it is important that clinicians assess the underlying reasoning for enhanced symptoms in order to guide management for optimal recovery. Lastly, no significant effects were found amongst any of the investigated variables on m-BESS scores.

In light of the increased awareness of the potential consequences of concussion, in addition to the acknowledged contribution of early detection, it is essential that the factors outlined in this study are reported during baseline SCAT3 testing and are considered post-concussion to direct objective management.
Chapter 4
Clinical utility of the SCAT3

4.1 Introduction

There are currently no stand-alone, objective measures available to identify concussion, however a multifaceted approach for the evaluation of concussion has been recommended by expert consensus (Guskiewicz et al., 2004; McCrory et al., 2005, 2013). The development of the sideline SCAT3 epitomizes such suggestions by encompassing measures of neurocognitive functioning, symptom endorsement and severity, and postural stability.

In order for a concussion assessment tool to be clinically useful, the test must be sensitive in that it has a high probability of accurately identifying a concussed athletes as having sustained a concussion, as well as specific in that it has a high likelihood of accurately identifying a healthy athlete as not having had a concussion. Research identifying the sensitivity and specificity of a diagnostic tool is integral to the test’s clinical utility, such that the magnitude of the results may be used to inform clinical practice and assist clinicians in making clinically relevant decisions.

Although the SCAT3 was developed as a sideline tool to diagnose concussion, due to the brevity of the test it has the potential to be implemented within clinical programs across the sequelae of concussion to provide meaningful information on an athlete’s gross cognitive functioning, balance, and symptomology in order to facilitate safe return-to-play decisions. The ability for a test to accurately distinguish between a healthy and concussed athlete is imperative in guiding clinical decision-making. For instance, if an athlete is identified as being healthy and is returned to play, but is continuing to suffer from concussive symptoms, the consequences are large in that the athlete’s symptom might exacerbate or a second concussive injury might be sustained, resulting in the risk for prolonged recovery and potential death (Bey & Ostick, 2009). Evidently, the clinical utility of a concussion test is of paramount importance.

Although the clinical utility of the SCAT3 in its entirety has yet to be explored empirically, the subtests included in the SCAT3 have been utilized extensively across North America, with the psychometric properties of such tests being well established.
The Post Concussion Symptom Scale

The literature consistently demonstrates the utility of graded symptom scales in facilitating the identification of concussion; however it has been suggested that the development of such tests has surpassed research examining their psychometric properties (Alla, Sullivan, Hale, & McCrory, 2009). For research that does exist, findings should be interpreted with caution, as many iterations of the same scale exist. For instance, Schatz and colleagues (2006) reported on the diagnostic utility of the ImPACT-PCSS, revealing a sensitivity of 81.9%, and specificity of 89.4% within 72 hours following concussion. The PCSS included in the ImPACT is similar to that included in the SCAT3, however there are slight differences that should be considered. As this study combined the utility of both the neurocognitive and symptom tests, it is difficult to ascertain the independent influence of the PCSS specifically.

Another study investigating the sensitivity and specificity of the ImPACT-PCSS found a sensitivity of 40.81% and specificity of 79.31% (Lau, Collins, & Lovell, 2011). Although this study assessed the diagnostic utility of the PCSS independently from neurocognitive performance, the sensitivity and specificity represented the ability to detect protracted recovery rather than acute diagnosis. Given that no known studies have assessed the sensitivity and specificity of the PCSS included in the SCAT3 specifically, it is of paramount importance that this is evaluated to validate its use in the sport context.

Standardized Assessment of Concussion

Several studies have reported post concussion deficits on the SAC in comparison to both baseline scores and to healthy control athletes (McCrea et al., 1998, 2003; McCrea, Kelly, Kluge, Ackley, & Randolph, 1997). However, the sensitivity and specificity of the SAC in identifying concussion has been sparsely researched.

Barr and McCrea (2001) were the first to report that immediately following concussion in high school and collegiate football athletes, a 1-point reduction on the SAC from baseline resulted in 94% sensitivity and 76% specificity in classifying concussed and non-concussed athletes. Equivalent results were reported in a study by McCrea (2001), finding the SAC to be 95% sensitive and 76% specific in identifying concussion on the sideline following a 1-point drop
from baseline pre-injury scores. McCrea and colleagues (2005) later found that the SAC was 80% sensitive and 91% specific in identifying concussion on the sideline in a sample of collegiate football players using a regression-based approach.

Although these findings suggest that the SAC is sensitive in identifying concussion, there is limited research evaluating the validity of this tool in other athletic populations, restricting the generalizability of such results. Additionally, a recent study has questioned the validity of the SAC in measuring baseline values, finding a ceiling effect for 63-70% of the items (Ragan et al., 2009). This may pose issues in the accuracy of comparing baseline values to post-injury scores using the SAC when deficits are slight, especially as time post-injury extends. Indeed, McCrea and colleagues (2005) demonstrated that the sensitivity of the SAC reduced from 80% to 2% between evaluations on the day of injury and seven days post-concussion.

**Balance Error Scoring System**

Previous research has revealed that the BESS is useful in detecting balance deficits immediately following concussion, however has reduced utility in tracking symptoms 3 days post-injury (Guskiewicz, 2001; Riemann, Guskiewicz, & Shields, 1999). Studies investigating the validity of the BESS revealed a sensitivity of 34-60% and specificity of 82-91% in identifying concussion on the sideline (Furman et al., 2013; McCrea et al., 2005). However, when used in conjunction with the SAC and a graded symptom scale, sensitivity of the BESS in identifying concussion increased (Giza et al., 2013). This suggests that although balance deficits are apparent following injury, the BESS is only moderately sensitive to detecting concussion and should be used with other measures to enhance its utility. Furthermore, as most research has investigated the original BESS, more research is needed to better determine the psychometric properties of the modified version.

**SCAT2**

A few studies have assessed the utility of the SCAT2 in identifying concussion. Mayfield and colleagues (2013) investigated the effect of concussion on SCAT2 total and component scores, finding significantly more symptoms and significantly worse balance and cognitive functioning on the sideline compared to baseline performance, with performance plateauing across 3 and 10 days post-injury. Additionally, Putukian and colleagues (2015) assessed the validity of the
SCAT2 total score within 24 hours post-injury, revealing that a 3.5 reduction in the total score from the athlete’s pre-injury baseline score was 96% sensitive and 83% specific in identifying athletes with concussion. This study also examined the utility of the SCAT2 total score when baseline data was not available, finding that a cut-point of 74.5 yielded the test to be 83% sensitive and 91% specific in identifying concussion within 24 hours post-injury. As the total score is no longer included in the revised SCAT3, research is needed to establish the sensitivity and specificity of the component scores independently and as a global measure across time.

The Automated Neuropsychological Assessment Metrics Sports Medicine Battery

In order to speak to the clinical utility of the SCAT3, it is critical that it be compared against another concussion assessment tool. The Automated Neuropsychological Assessment Metrics Sports Medicine Battery (ANAM) is a library of such computerized batteries designed to measure neurocognitive functioning across serial administrations (Kabat, Kane, Jefferson, & DiPino, 2001). Originally developed in sponsorship with the U.S Military Department of Defense for assessing cognitive, perceptual, and psychomotor performance in concussed militants, ANAM has since been expanded to be utilized in the context of sport medicine (Reeves, Winter, Blieberg, & Kane, 2007).

Coldren and colleagues (2012) evaluated the utility of the ANAM battery as a diagnostic tool for detecting neurocognitive deficits in concussed soldiers, finding that within the first 72 hours following injury, concussed soldiers demonstrated impairments on the ANAM compared to controls, however at more than 10 days post-injury the ANAM appeared to lack any clinical utility. Similarly, Register-Mihalik and colleagues (2013) assessed the utility of the ANAM, SOT and GCS in a sample of collegiate football players within five days of sustaining a concussion, finding the ANAM to have little sensitivity in detecting neurocognitive deficits, however increased when the tests were combined. As not all of the modalities in the ANAM battery were included in these research designs, further research is required to assess the utility of the sports medicine battery specifically to compare against the SCAT3.

4.2 Rationale
A review of the literature on the utility of the SCAT3 components in identifying concussion has revealed many knowledge gaps. Firstly, although it is evident that many studies exist identifying the validity of the major SCAT3 component scores, there is limited research defining the sensitivity and specificity of the SCAT3 as a collective measure across time. Though many of the studies have assessed the validity of such tools in the acute phase of injury, there is value in determining how sensitive the SCAT3 is over the concussion sequelae to inform clinical decision-making.

Additional gaps in knowledge pertain to the methodology used in previous research. Specifically, the utility of the SAC has only been assessed in male football players without looking at additional athletic cohorts, the PCSS has only been evaluated in combination with ImPACT and in a slightly different iteration making the independent sensitivity of the PCSS unknown, and finally no studies have assessed the utility of the modified BESS, only assessing the original version.

Moreover, although research findings on the clinical utility of the SCAT2 total score with and without baseline data are informative, the revised SCAT3 has been modified to exclude this score, warranting the need for this research to be updated (Putukian et al., 2014). Baseline testing has previously been advised due to differences in scores being found between healthy athletes on a range of demographics that may affect post-concussion test interpretation (Valovich McLeod et al., 2012), however experts do not currently endorse baseline testing as mandatory (McCrory et al., 2013). With the large time and financial commitment associated with baseline testing, it is imperative to determine how the utility of the SCAT3 differs when an athlete’s baseline scores are or are not available to inform future clinical practice standards.

Lastly, there is no previous research comparing the utility of the SCAT3 to other concussion assessment tools. By comparing the utility of the SCAT3 to a computerized concussion assessment battery, the clinician’s understanding of the optimal time to implement each tool with be enhanced.

4.3. Objectives
Given the rationale for the current study, the objective was to evaluate the clinical utility of the SCAT3 by:

a. Defining the sensitivity and specificity of the measures’ subtests independently and in combination across time;
b. Comparing the sensitivity and specificity of the SCAT3 to a computerized neurocognitive concussion assessment tool;
c. Investigating differences in the sensitivity and specificity of the SCAT3 in the presence and absence of baseline data

4.4. Methods and procedure

4.4.1. Design
The study was conducted using a prospective cohort design, with baseline and serial repeat testing.

4.4.2. Participants
A total of 524 varsity athletes participating in a high-risk sport at the University of Toronto between August of 2014 and December of 2015 participated in baseline testing. A total of 23 concussed athletes (10 men) and 22 control athletes (9 men) participated in serial repeat testing. Participants were recruited from the University of Toronto’s varsity athletics teams, including men and women’s ice hockey, lacrosse, volleyball, soccer, rugby, and basketball, men’s football, baseball, and wrestling, and women’s field hockey.

Athletes were eligible to participate in the study if they were able to speak and understand English fluently, if they provided informed consent, and if they had no history of neurological or balance disorders or a previous concussion within the last 6 months. Prior to participation in this study, all athletes provided informed consent as approved by the Health Sciences Research Ethics Board (REB) at the University of Toronto. The informed consent documents taken from the REB application can be found in Appendix 1.

4.4.3. Outcome measures
The main outcome measures were comprised of the major SCAT3 and ANAM subtests.
SCAT3
The main subtests measured in the SCAT3 included: (1) Post Concussion Symptom Scale total symptom and symptom severity scores, (2) SAC, and (3), m-BESS. For a detailed description of these measures, please refer to the outcome measures section of Chapter 3.

Automated Neuropsychological Assessment Metrics Sports Medicine Battery
The ANAM4™ is a library of computerized neurocognitive test batteries, with 8 main modules being of interest in the current study:

1. Code Substitution (CDS) is comparable to the symbol-digit test used in traditional paper-and-pencil neuropsychological test batteries used to assess visual search, working memory, and sustained attention.
2. The Procedural Reaction Time (PRO) is a continuous reaction-time test that assesses processing speed, visuo-motor reaction time, and attention.
3. Spatial Processing (SPD) is a test that examines visual spatial skills.
4. Memory Search (ST6) is a test that examines verbal working memory, immediate recognition, and attention.
5. Code Substitution Delayed (CDD) is the same as CDS, but is administered at the end of the battery as a measure of learning and delayed visual recognition memory.
6. Simple Reaction Time (SRT) is a test that measures visuo-motor processing speed, simple motor speed, and attention by requiring the participant to press a button as quickly as possible following a stimulus prompt.
7. Simple Reaction Time Repeated (SR2) is identical to SRT, but administered at a later time during the testing session.
8. Matching to Sample (M2S) is a measure of visual-spatial processing, working memory, and visual recognition memory.

Data are collected for each module and a number of different summary variables are produced. The variables selected for analysis in this study included mean reaction time for correct responses (MeanRTCorr), percent correct (PercCorr), and throughput, which is a composite score of speed and accuracy.
4.4.4. Procedure and data collection

All testing was completed in the concussion laboratory at the University of Toronto. The ANAM was administered on a desktop computer with the participant using a mouse to record his/her responses, taking approximately 30 minutes to complete. The SCAT3 was administered by trained research assistants or graduate students and took approximately 10 minutes to complete. Athletes also underwent blood sampling and vision testing as a part of a larger scale study, making the total testing time approximately one hour.

The SCAT3 and the ANAM were administered to all athletes prior to the commencement of the athletic season in order to attain a baseline measure. Baseline testing was completed in one setting with up to eight athletes from the same team, rotating amongst test administrators.

The team’s athletic therapist notified the research coordinator when an athlete was suspected of sustaining a concussion and were recruited to participate in the study via phone or email. Athletes were confirmed to have had sustained a concussion by a medical physician and were recruited to complete the SCAT3 and ANAM within the first 3 to 5 days following injury and then again at three weeks post-injury unless they were medically cleared prior to this time, where athletes completed testing on the date of medical clearance. There were 2 athletes medically cleared upon the first visit to the lab and a total of 10 athletes medically cleared upon the second visit to the lab. A few concussed athletes recruited in the 2014-2015 season only completed the PCSS portion of the SCAT3 during serial testing, however the remaining athletes completed the test in its entirety.

After identifying an athlete as having sustained a concussion, all efforts were made to match a healthy control athlete on the characteristics of sex, age, sport, and history of concussions using the baseline demographic information. Efforts were also made to recruit control participants close in proximity to the time of the concussed athletes in order to standardize academic and varsity pressures. Control athletes completed the equivalent testing design. A visual representation of the testing timeline is presented in Figure 4.1 below.
4.4.5. Data analysis

I. Post-concussion analyses

All analyses were performed using IBM SPSS Version 23. To investigate whether there were any differences between the concussed and control athletes’ demographics, testing schedules, or medical history, independent sample t-tests and chi square tests were performed.

Tests of normality revealed that the SCAT3 and ANAM subtest scores were non-normally distributed, and therefore the appropriate nonparametric data analyses were performed. As there is no known non-parametric test equivalent to a mixed-factor ANOVA, separate between- and within-group analyses were performed.

In order to determine whether there were any significant differences between the control and concussed athletes on each of the SCAT3 and ANAM subtests at a specific point in time, separate Mann-Whitney U analyses were performed. Mann-Whitney U tests were also performed to determine if there were sex differences between the control and concussed athletes. Levene’s test of homogeneity was performed to determine whether scores varied significantly across
groups. If variances differed, mean ranks were compared for significance, however if the variances were similar, median scores were compared for significance.

In order to determine whether there were any significant differences within each group on the SCAT3 and ANAM subtests across time, separate Friedman tests were performed. If a significant main effect was found on the Friedman test, post hoc pairwise analyses were performed with Dunn-Bonferroni adjustments. The effect of sex was also assessed within the concussed and control athletes across time by performing the Friedman analysis. Mauchly’s sphericity tests were performed for all measures to determine whether the variances in scores significantly differed across time.

II. Sensitivity and specificity analyses
Analyses were performed using IBM SPSS Version 23. In order to determine the clinical utility of the SCAT3 and ANAM across time, the sensitivity and specificity was evaluated using two different approaches.

1. The first method utilized a test-retest approach by applying the Iverson adapted formula of Jacobson and Truax to construct reliable change indices (RCI) (Iverson, Lovell, & Collins 2003).
Specifically, Pearson correlation coefficients were calculated for the SCAT3 and ANAM subtests between baseline and re-test using the healthy control data. The standard deviations (SD) and reliability coefficients (r) were then used to compute the standard error of the means using the formulas:

\[ SEM_1 = SD_0 \sqrt{1 - r_1} \]  
(Standard deviation from baseline multiplied by the square root of 1 minus the test-retest coefficient between baseline and time 1).

\[ SEM_2 = SD_1 \sqrt{1 - r_1} \]  
(Standard deviation from time 1 multiplied by the square root of 1 minus the test-retest coefficient between baseline and time 1).
The same formulas were used to compute the SEMs for the 3-week time point by substituting the appropriate standard deviation and reliability values.

The standard error of the difference (Sdiff) was then calculated as:

\[ S_{diff} = \sqrt{SEM_1^2 + SEM_2^2} \]

(Square root of the sum of the squared SEMs for each testing occasion)

This methodology for computing the Sdiff uses the SEM for baseline and retest in order to correct for potential practice effects, rather than estimating the Sdiff by multiplying the squared baseline SEM by two (Iverson, Lovell, & Collins, 2003).

The RCIs were then calculated for the .70, .80, and .90 confidence intervals (CI) by multiplying the Sdiff by the corresponding z values (i.e. a CI of .90 corresponds to a z value of 1.645, a CI of .80 = z of 1.28, and CI of .70 = z of 1.4). Participants were then classified as either “concussed” or “healthy” using the RCIs of each test and at each time point as a cut-off score. Therefore, the .90 confidence interval creates the most stringent cut-off score for classifying an athlete as having a concussion as it corresponds to the largest z value, whereas the .70 confidence interval is the least stringent.

The number of true positive, true negatives, false positive, and false negatives were then counted, where true positives are correctly classifying a concussed athlete as being abnormal, true negatives are correctly classifying a healthy athlete as being normal, false positives are misclassifying an athlete as having a concussion when they are actually healthy, and false negatives are misclassifying a healthy athlete as being abnormal. The number of true positives naturally decreased as time post-injury extended as more athletes became medically cleared, whereas the number of true negatives increased. Sensitivity was computed as the number of true positives divided by the sum of true positives and false negatives, whereas specificity was calculated as the number of true negatives divided by the sum of true negatives and false positives.

The sensitivity of the combined SCAT3 was found by using the following formula:

\[ \text{Sensitivity combined} = 1 - (1 - TS_{SEi}) \times (1 - SS_{SEi}) \times (1 - mBESS_{SEi}) \times (1 - SAC_{SEi}) \]
(1 minus 1 minus the sensitivity of the total symptom score multiplied by 1 minus the sensitivity of the symptom severity score multiplied by 1 minus the sensitivity of the m-BESS score multiplied by 1 minus the sensitivity of the SAC score) (Parikh et al., 2008).

The specificity of the combined SCAT3 was found using the same formula, while substituting for each independent test’s specificity value. This formula was also used to find the combined sensitivity and specificity values for each of the ANAM summary data scores, namely the throughput score, reaction time, and percent correct items, as well as a combination of these measures.

2. The second methodology for examining the clinical utility of the SCAT3 used a normative approach.

For the SCAT3, norms were taken from the 524 athletes’ baseline data, whereas for the ANAM, norms were taken from the ANAM Sport-College Reference Values that comprises of data collected from the University of North Carolina, University of Oklahoma, and the University of Toronto athletic programs (range of 494-1124 athletes across ANAM subtests). Normative data were also separated into male and female norms, resulting in 334 males and 190 females for the SCAT3 and a range of 186-411 females and 306-715 males for the ANAM. Normative data was used to create cut-off scores below the 1st, 5th, 10th, and 15th percentiles, with the most stringent cut-off corresponding to the 1st percentile and the least stringent to the 15th percentile. Athletes were once again classified as being “concussed” or “healthy” by comparing the re-test scores to the cut-off scores. Sensitivity and specificity were then computed by counting the number of true positives, true negatives, false positives, and false negatives. The combined sensitivity and specificity of the SCAT3 was found using the same formula as previously presented.

III. ROC Curve Analyses
Lastly, receiver operating characteristic (ROC) curves were created using IBM SPSS Version 23 to establish the most appropriate cut-off points set at a 95% confidence interval for each SCAT3 subtests using a baseline-retest approach by comparing the athlete’s own baseline score to his/her
retest score and a normative approach using the means and medians of the baseline data as a comparator to the athlete’s retest score.

MedCalc™ software was then used to statistically compare the area under the ROC curves between the normative and baseline-retest approach (Delong & Carolina, 2016).

4.5. Results
4.5.1 Demographics
A total of 23 athletes (10 men) having sustained a concussion across the 2014-2015 or 2015-2016 athletic season participated in this study. On average, athletes were 20.04 ± 1.99 years of age. The mean time between baseline testing and the first time tested post-injury was 205.19 ± 207.7 days. The mean time between the first and second time tested post-injury was 17.05 ± 7 days. At the first time of testing post-injury, 2 female athletes were medically cleared to return-to-play and at the second time of testing post-injury, an additional 10 athletes were medically cleared to return-to-play, of which 6 were men. Of the newly concussed athletes, 9 reported having had no previous concussions, 11 had one to two previous concussions, and two athletes had three or more concussions. Moreover, one athlete reported being previously diagnosed with a LD/ADHD, one athlete was previously diagnosed with migraine headaches, and five athletes reported having a family member previously diagnosed with a mental health disorder. The concussed athletes were participating in men’s (n= 2) and women’s (n=4) rugby, football (n= 7), men’s (n=1) and women’s (n=3) volleyball, men’s (n=2) and women’s (n= 1) ice hockey, women’s basketball (n= 3), women’s lacrosse (n= 2), and men’s soccer (n= 1).

Twenty-two healthy control athletes (9 men) were recruited to participate in this study. On average, athletes were 19.45 ± 1.92 years of age. The mean time between baseline testing and the second visit was 301 ± 175.69 days. The mean time between the first and second time tested was 19.57 ± 3.92 days. Of the healthy control athletes, 11 reported having had no previous concussions, 10 had one to two previous concussions, and one athlete had three or more concussions. Moreover, two athletes reported being previously diagnosed with a LD/ADHD, eight athletes reported having a family member previous diagnosed with a mental health disorder, and five athletes reported having a personal mental health history. The control athletes
were participating in men’s (n= 1) and women’s (n=4) rugby, football (n= 7), men’s (n=1) and 
women’s (n=3) volleyball, men’s (n=2) and women’s (n= 1) ice hockey, women’s basketball (n= 
3), women’s lacrosse (n= 2), and men’s soccer (n= 1). There were no significant differences 
between the concussed and control athletes’ demographics, testing schedules, or medical history, 
however there were more controls athletes with a history of mental health compared to 
concussed athletes, with this approaching significance, $\chi^2(1)= 3.805$, p = .0511.

4.5.2 Comparison of concussed and control Athletes: SCAT3

*Between groups*

Levene’s test of homogeneity revealed that the variances in SCAT3 component scores were 
significantly different between groups; therefore mean ranks were compared for significance for 
all of the results presented.

*Concussion versus Controls*

Mann Whitney U tests run between control and concussed athletes at the acute phase of injury 
revealed a significant main effect on the PCSS total symptom (U= 405, z= 3.467, p= .001) and 
symptom severity (U= 382, z= 2.944, p=.003) scores, with concussed athletes reporting more 
symptoms (mdn= 8, mean rank= 29.61) of greater severity (mdn= 9, mean rank= 28.61) 
compared to controls (TS: med= 1.5, mean rank= 16.09; SS: mdn= 3, mean rank= 17.14). The 
SAC and m-BESS were not statistically significant between groups at the acute phase of injury. 
Moreover, none of the SCAT3 component scores were significantly different between the 
concussed and control athletes at baseline and 3 weeks post-injury.
Figure 4.2. Mean ranking of PCSS total symptom score significantly greater in concussed athletes at acute phase of injury compared to controls.

Figure 4.3. Mean ranking of PCSS symptom severity score significantly greater in concussed athletes at acute phase of injury compared to controls.

Effect of Sex on SCAT3 Outcomes Post-Concussion

There were no significant differences on the SCAT3 component scores between males or females within the concussed or controls athletes across all testing times as revealed by Mann Whitney U tests. However, it was found that within the concussed group, females reported more
symptoms (M= 10.23, SD= 30.77) of greater severity (M= 26.3, SD= 31.06), compared to males (TS: M= 8, SD= 5.68, SS: M= 12.2, SD= 14.63) at the acute phase of injury.

When comparing for differences on the SCAT3 between the concussed and control groups within the female participants, no significant findings were revealed at baseline or at the third time tested. However, at the acute time of testing, a significant difference was found between groups on the total symptom score (U= 125, z= 2.085, p= .039) with concussed athletes reporting more symptoms (mdn= 8, mean rank= 16.62) compared to controls (mdn= 3, mean rank= 10.38). There was also a trend for greater symptom severity in concussed athletes (mdn= 11, mean rank= 16.42) at this time compared to controls (mdn= 4, mean rank= 10.58), however this did not reach significance (U= 122.5, z= 1.957, p= .05). There were no significant differences found between concussed and control athletes within the female participants on the SAC or m-BESS.

When comparing for differences on the SCAT3 between the concussed and control groups within the male participants, Mann Whitney U tests revealed a significant effect on the m-BESS at baseline (U= 18.5, z= -2.207, p= .027), with concussed male athletes performing significantly worse (mdn= 26, mean rank= 7.35) compared to healthy male controls (mdn= 28, mean rank= 12.94). No other significant differences were found between groups at baseline or 3 weeks post-injury.

At the acute phase of injury it was revealed that males with a concussion reported significantly more symptoms (U= 83, z= 3.129, p= .001; mdn= 6.5, mean rank= 13.8) of greater severity (U= 76, z= 2.552, p= .01; mdn= 7.5, mean rank= 13.1) compared to controls (TS: mdn= 1, mean rank= 5.78; SS: mdn= 1, mean rank= 6.56). No significant differences between groups were found on the SAC or m-BESS at this testing time.

Within Groups
Mauchly’s test of sphericity revealed that the variances in the SCAT3 outcome measures were significantly different across time. Therefore mean ranks were compared for significance for all of the results presented.

Controls
Friedman tests run on the control group across time revealed a significant main effect for the total symptom ($\chi^2(2)= 6.371, p =.041$) and symptom severity ($\chi^2(2)= 7.497, p =.024$) scores. Post hoc pairwise analyses revealed no significant differences between the times tested, however a trend was observed such that the control athletes reported lower symptoms of lesser severity over time. There were no significant differences between the SAC or m-BESS across time in the control group.

Figure 4.4. Main effect of time within control athletes on PCSS total symptom score.
Figure 4.5. Main effect of time within control athletes on PCSS symptom severity score.

Within the female control athletes, the Friedman analysis revealed a significant main effect for the SCAT3 total symptoms score, $\chi^2(2)= 7.409, p= .025$, with post hoc analyses demonstrating significantly lower reporting of symptoms at the third (mdn= 2, mean rank= 1.42) compared to second time tested (mdn= 3, mean rank= 2.46), $p= .032$). No other significant differences were found across time on the symptom severity, SAC, or m-BESS scores. Within the male control athletes, there were no significant differences on any of the SCAT3 component scores across time, however, once again there was a trend for lower symptom reporting over time.

Concussions
Friedman tests revealed a significant difference between total symptom ($\chi^2(2)= 17.452, p =.000$) and symptom severity ($\chi^2(2)= 16.692, p =.000$) scores across time in the concussed group of athletes. Post hoc pairwise analyses revealed that symptom severity scores were significantly greater at the acute phase of injury (mdn= 9, mean rank= 2.62) compared to 3 weeks post-injury (mdn= 3, rank= 1.35), $p=.000$. Although the difference between symptom severity scores at baseline and the acute phase of injury was not significant, there was a trend for higher symptom severity reporting at the time post-injury ($p= .098$). Post hoc analyses further revealed that the total symptoms score was significantly greater acutely (mdn= 8, mean rank= 2.68) compared to
at baseline (mdn= 3 mean rank= 1.9), p= .043 and at 3 weeks post-injury (mdn= 2.5, mean rank= 1.42), p= .000. There were no significant differences across time on either the SAC or m-BESS in the concussed athletes.

**Figure 4.6.** Mean ranking of PCSS total symptom score in concussed athletes significantly greater at acute phase of injury compared to baseline and 3 weeks post-injury.

**Figure 4.7.** Mean ranking of PCSS symptom severity score in concussed athletes significantly greater at acute phase of injury compared to 3 weeks post-injury.
Female Concussions
Friedman tests within the female concussed athletes revealed a significant main effect of time for the total symptom ($\chi^2(2)= 6.4, p = .041$) and symptom severity ($\chi^2(2)= 8.667, p = .013$) scores, with post hoc tests revealing significantly greater symptom severity scores at the acute phase of injury (mdn= 11, mean rank= 2.65) compared to 3 weeks post-injury (mdn= 2.5, mean rank= 1.35), $p= .011$. Follow-up analysis of the total symptom main effect revealed no significant differences between the times tested, however there was a trend of heightened symptoms at the acute phase of injury, compared to baseline and 3 weeks post-injury.

Male Concussions
Friedman tests within the male concussed athletes revealed a significant main effect of time for the total symptom ($\chi^2(2)= 11.842, p = .003$) and symptom severity ($\chi^2(2)= 8.051, p = .018$) scores, with post hoc tests revealing significantly heightened symptoms of greater severity at the acute phase of injury (TS: mdn= 6.5, mean rank= 2.75; SS: mdn= 7.5, mean rank= 2.6) compared to 3 weeks post-injury (TS: mdn= 2.5, mean rank= 1.25; p= .002 SS: mdn= 3, mean rank= 1.35, $p= .016$).

4.5.3 Comparison of concussed and control athletes: ANAM

Between Groups
Levene’s test of homogeneity revealed that the variances in ANAM scores were significantly different between groups. Therefore mean ranks were compared for significance for all of the results presented.

Between Concussions and Controls
Mann Whitney U tests revealed no significant differences between groups at baseline for any of the outcomes measures for the following tasks: Matching to Sample (M2S), Simple Reaction Time (SRT), Simple Reaction Time Repeated (SR2), Procedural Reaction Time (PRO), and Memory Search (ST6) tasks. However, there was a significant difference between control and concussed athletes at baseline on the Spatial Processing (SPD) task ($U= 84.5, z= -3.067, p= .002$), such that control athletes (mdn=95, rank= 25.28) had significantly more percent correct compared to concussed athletes (mdn= 90, rank= 14.45). Throughput scores and mean reaction time for correct responses did not significantly differ between groups on the SPD task.
At the acute phase of injury, Mann Whitney U tests revealed no significant differences between groups on any of the ANAM subtests’ outcome measures. Although not significant, it was found that concussed athletes performed more poorly on the CDD test, as demonstrated by lower throughput scores (mdn= 56.69; mean rank= 18.84, p= .059) and higher reaction time for correct responses (mdn= 983.19, mean rank= 26.18, p= .057), compared to controls (TP: mdn= 61.85, mean rank= 26.16; RTcorr: mdn= 921.22, mean rank= 18.82).

At the third time point of testing, Mann Whitney U tests revealed no significant differences between any of the outcome measures on the Code Substitution (CDS), Code Substitution Delayed (CDD), Mathematical Processing (MTH), Simple Reaction Time (SRT), Simple Reaction Time Repeated (SR2), Procedural Reaction Time (PRO), and Memory Search (ST6) tasks. However, a significant difference was found between concussed and control athletes on the SPD task (U= 295, z= 1.973, p= .048), with post hoc analyses revealing that concussed athletes had significantly greater percent correct items (mdn= 100, mean rank= 25.05), compared to controls (95, mean rank= 17.95).

Between Sexes Baseline
When comparing for differences in ANAM test responses across sexes at baseline, Mann Whitney U tests revealed that all but the CDD test were insignificant, with female athletes having heightened throughput scores (mdn= 60.28, mean rank= 24.17; U= 280, z= 2.741, p= .005) and faster reaction times (mdn= 967.09, mean rank= 16.35; U= 100, z= -2.398, p= .016) compared to male athletes (TP: mdn= 48.46, mean rank= 14; RT: mdn= 1106.75, mean rank= 25.25).

Between Sexes within Concussed Athletes
Within the concussed set of athletes, no significant differences were found between sexes on any of the ANAM subtest outcome measures at baseline or at the acute phase of injury. However, at 3 weeks post-injury, the M2S throughput score was found to be significantly different between sexes (U= 24, z= -2.132, p= .034), with male concussed athletes performing significantly better (mdn= 54.89, mean rank= 14.33) compared to females (mdn= 36.735, mean rank= 8.5). It should be noted however that at 3 weeks post-injury only 4 of the 12 female concussed athletes were
medically cleared, whereas 5 of the 9 male concussed athletes were cleared at this time. There were no other significant differences found between sexes at the third time of testing.

**Between Sexes within Control Athletes**

There were no significant differences between sexes within the control athletes on any of the ANAM subtests’ outcome measures at baseline or at the third time of testing. However, at the second time of testing, the M2S percent correct significantly differed between sexes (U= 25.5, z= -2.313, p= .025) with males performing significantly better (mdn= 100, mean rank= 15.17) compared to females (mdn= 95, mean rank= 8.96). No other significant differences between sexes was found on any of the ANAM subtest outcome measures at the second time of testing.

**Between Concussed and Control Athletes within Females**

Within the female participants, Mann Whitney U tests run between concussed and control athletes revealed that at baseline, SPD percent correct was significantly different between groups, U= 24, z= -2.731, p=.006, with controls (mdn= 37.08, mean rank= 15.5), performing significantly better than athletes whom later sustained a concussion (mdn= 35.48, mean rank= 8.18). No other significant differences were found between athletes at baseline or at 3 weeks post-injury.

At the acute phase of injury within female participants, Mann Whitney U tests revealed a significant difference between concussed and control athletes on the PRO and SRT tests. Specifically, the PRO throughput score (U= 36, z= -2.284, p= .022) and reaction time for correct responses (U= 118, z= 2.176, p=.03) significantly differed between groups, with female concussed athletes (TP: mdn= 92.33, mean rank= 9.5; RT: mdn= 635.185, mean rank= 16.3) performing significantly worse compared to female control athletes (TP: (mdn= 103.72, mean rank= 16.23; RT: mdn= 543.44, mean rank= 9.92). Additionally, the SRT throughput score (U= 29.5, z= -2.639, p=.007) and reaction times (U= 127, z= 2.665, p=.007) were significantly different between groups, with concussed athletes (TP: mdn= 192.935, mean rank= 8.96; RT: mdn= 311, mean rank= 17.08) performing significantly worse than the healthy control athletes (TP: mdn= 207.61, mean rank= 16.73; RT: mdn= 289, mean rank= 9.23). No other significant differences were found between groups on any of the ANAM subtest outcome measures at the acute phase of injury.
Between Concussed and Control Athletes within Males

Within the male participants, Mann Whitney U tests revealed a significant difference between the control and later-concussed athletes at baseline on the SR2 throughput score (U= 51, z= 1.995, p= .046) and reaction time (U= 13, z= -1.995, p= .046), with control athletes (TP: mdn= 199.14, mean rank= 6.12; RT: mdn= 301.31, mean rank= 10.88) performing significantly worse than athletes whom later sustained a concussion (TP: mdn= 228.39, mean rank= 10.88; RT: mdn= 295, mean rank= 6.12). There were no other significant differences between male concussed or control athletes at baseline, at the acute phase of injury, or at 3 weeks post-injury.

Within Groups

Mauchly’s test of sphericity revealed that the variances in the ANAM outcome measures were significantly different across time. Therefore mean ranks were compared for significance for all of the results presented.

Control Athletes

Friedman tests revealed no significant differences within the control group across time on the CDS, CDD, PRO, SPD, SRT, and SR2 tasks for any of the outcome measures. However, the ST6 task was found to be significantly different in the control group across time ($\chi^2(2)= 6.737, p= .034$), with post hoc tests revealing significantly greater throughput scores at the second time tested (mdn= 84.14, mean rank= 2.42), compared to baseline (mdn= 75.76, mean rank= 1.58), $p= .028$. The reaction time for correct responses in the ST6 task was also found to significantly differ across time in the control group of athletes ($\chi^2(2)= 7.053, p= .029$), with reaction times significantly improving at the second time of testing (mdn= 705.53, mean rank= 1.53) compared to baseline (mdn= 758.64, mean rank= 2.37), $p= .028$. Additionally, the M2S throughput score was found to be significantly different within the control athletes over time ($\chi^2(2)= 6.737, p= .034$), with post hoc tests revealing significantly greater scores at the third time tested (mdn= 40.95, mean rank= 2.42), compared to baseline (mdn= 38.12, mean rank= 1.58), $p= .028$. Taken together, these results suggest a trend towards improvement in the control group, likely driven by practice effects.
Female Controls Over Time
Within the healthy female controls, Friedman tests revealed no significant differences on any of the ANAM subtest outcome measures across time. However, there was a trend for improved reaction time for correct items on the SPD, PRO, and ST6 tasks between baseline and subsequent test trials.

Male Controls Over Time
Within the male control athletes, Friedman tests revealed a significant main effect of time on the CDD percent correct measure, $\chi^2(2) = 9.172, p = .010$, with a significant post hoc finding of improved scores at the second time of testing (mdn= 97.22, mean rank= 2.56) compared to baseline ((mdn= 91.67, mean rank= 1.19). No other significant differences were found within the male control athletes over time.

Concussed Athletes
Within the concussed set of athletes, Friedman tests revealed that the PRO throughput score ($\chi^2(2)= 8, p=.018$) and reaction time ($\chi^2(2)= 6.125, p=.047$) significantly differed over time. Post hoc analyses of the main effect of time on the PRO throughput score revealed significantly greater performance at 3 weeks post-injury (mdn= 102.63, mean rank= 2.5) compared to when acutely injured (mdn= 95.36, mean rank= 1.5), $p=.014$. Post hoc analyses for the PRO reaction time main effect revealed no significant differences across time, however median reaction time was lower 3 weeks post-injury (mdn= 570.78, mean rank= 1.69) and at baseline (mdn= 563.06, mean rank= 1.81), compared to when acutely concussed (mdn= 618.25, mean rank= 2.5).

A significant main effect of time was also revealed for the SR2 throughput scores ($\chi^2(2)= 8, p=.018$) and reaction times ($\chi^2(2)= 8, p=.018$), as well as for the SRT throughput scores ($\chi^2(2)= 9.125, p= .01$) and reaction times ($\chi^2(2)= 9.125, p= .01$). Post hoc analyses revealed that within the concussed athletes, performance was significantly better and faster at baseline (SR2 TP: mdn= 208.74, mean rank= 2.5; RT: mdn= 287.43; mean rank= 1.5; SRT TP: mdn= 204.31, mean rank= 2.5; RT: mdn= 293.68, mean rank= 1.5) compared to when acutely concussed (SR2 TP: mdn= 196.41, mean rank= 1.5, $p=.014$; RT: mdn= 305.51, mean rank= 2.5, $p=.014$; SRT TP: mdn= 195.03, mean rank= 1.44, $p=.008$; RT: mdn= 310.1, mean rank= 2.5, $p=.008$).
A significant main effect of time was also revealed on the M2S percent correct outcome measure ($\chi^2(2) = 9.918, p = .007$), with no significant post hoc findings, however a trend of improvement was found between baseline and the third time of testing.

Lastly, a significant main effect was found on the SPD throughput score ($\chi^2(2) = 10.5, p = .005$) and percent correct measure ($\chi^2(2) = 7.538, p = .023$). Post hoc analyses revealed a significant improvement in performance between baseline testing (TP: mdn= 32.89, mean rank= 1.5; PC: mdn= 90, mean rank= 1.56) and 3 weeks post-injury (TP: mdn = 42.54, mean rank= 2.62, p= .004; PC: mdn= 100, mean rank= 2.44, p= .04).

No significant differences were found on the CDS, CDD, or ST6 tasks over time within the concussed athletes.

**Concussed Females Over Time**

Within the female concussed athletes, Friedman tests revealed a significant main effect of time on the PRO throughput score, $\chi^2(2) = 6.889, p = .032$, with post hoc analyses revealing significantly worse performance when acutely concussed (mdn= 92.33, mean rank= 1.33) compared to 3 weeks post-injury (mdn= 100.765, mean rank= 2.56), p= .029. There was also a significant main effect of time on the PRO reaction time measure ($\chi^2(2) = 6.222, p = .045$), with no significant post hoc findings, however there was a trend of slower reaction time for correct responses at the acute phase of injury (mdn= 633.46, mean rank= 2.67) compared to 3 weeks post-injury (mdn= 577.575, mean rank= 1.78), p= .055. No other significant differences were found on any of the ANAM subtest measures over time in the female concussed athletes, however the SRT reaction times and throughput scores demonstrated poorer performance when acutely concussed, compared to baseline and 3 weeks post-injury.

**Concussed Males Over Time**

Within the male concussed athletes, Friedman tests revealed a significant main effect of time for the M2S percent correct measure ($\chi^2(2) = 8.818, p = .012$), with post hoc tests revealing a significant improvement in performance between baseline (mdn= 90, mean rank= 1.21) and 3 weeks post-injury (mdn= 100, mean rank= 2.57), p= .012. Additionally, the SPD throughput score was found to significantly differ across time ($\chi^2(2) = 8.857, p = .012$), with post hoc tests
revealing improved performance 3 weeks post-injury (mdn= 42.54, mean rank= 2.86), compared to baseline (mdn= 31.145, mean rank= 1.86), p= .010. No other significant differences were found on any of the ANAM subtests’ outcome measures within the male concussed athletes across time. However, the SR2 throughput and reaction time scores demonstrated a trend of poorer performance at the acute phase of injury compared to baseline.

4.5.4 Sensitivity and specificity of the SCAT3

The reliability values for the SCAT3 scores based on the healthy control data are presented in table 4.1. For a full record of the sensitivity and specificity results of the SCAT3 components independently and combined using a baseline retest approach across the .70, .80, and .90 confidence intervals and a normative approach including all participants, male participants only, and female participants only using the lowest 1\textsuperscript{st}, 5\textsuperscript{th}, 10\textsuperscript{th}, and 15\textsuperscript{th} percentiles, reference table 4.2 for the acute phase of injury, and table 4.3 for the post-acute phase of injury.

Table 4.1. Pearson correlation coefficient reliability values for SCAT3 scores.

<table>
<thead>
<tr>
<th>SCAT3</th>
<th>Baseline: 3-5 Days</th>
<th>Baseline: 3 Weeks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Symptoms</td>
<td>.407</td>
<td>.32</td>
</tr>
<tr>
<td>Symptom Severity</td>
<td>.296</td>
<td>.223</td>
</tr>
<tr>
<td>SAC</td>
<td>.468</td>
<td>.445</td>
</tr>
<tr>
<td>mBESS</td>
<td>.01</td>
<td>-.14</td>
</tr>
</tbody>
</table>

From these tables, there are a few key findings that must be highlighted. Firstly, it can be seen that the least stringent cut-off scores created using the .70 confidence interval in the retest approach and the lesser than 15\textsuperscript{th} percentile used in the normative approach led to the greatest sensitivity values across the two times tested; and although the specificity was slightly lower using these indices, they remained very high. These values will be discussed moving forward.

In both of the acute and post-acute phases of injury, the sensitivity was found to be highest for the total symptom score of the PCSS, followed by the symptom severity score, the m-BESS, and then the SAC. The highest sensitivity reached by the total symptom score was 72.7\% and 60\% for the symptom severity score using the female normative approach at the acute phase of injury. The combined sensitivity of the PCSS total symptom and symptom severity scores ranged from
Table 4.2 Sensitivity and specificity of SCAT3 components individually and combined at the acute phase of injury using a test-retest and normative approach.

<table>
<thead>
<tr>
<th>Combined SCAT3</th>
<th>m-BESS</th>
<th>SAC</th>
<th>Symptom Severity</th>
<th>Total Symptoms</th>
<th>Utility</th>
</tr>
</thead>
<tbody>
<tr>
<td>76.20%</td>
<td>15.4%</td>
<td>15.4%</td>
<td>36.8%</td>
<td>47.4%</td>
<td>Se</td>
</tr>
<tr>
<td>100%</td>
<td>91.3%</td>
<td>91.7%</td>
<td>100%</td>
<td>100%</td>
<td>Sp</td>
</tr>
<tr>
<td>60.6%</td>
<td>7.7%</td>
<td>0%</td>
<td>26.3%</td>
<td>42.1%</td>
<td>Se</td>
</tr>
<tr>
<td>100%</td>
<td>91.3%</td>
<td>91.7%</td>
<td>100%</td>
<td>100%</td>
<td>Sp</td>
</tr>
<tr>
<td>78.5%</td>
<td>7.14%</td>
<td>0%</td>
<td>39.1%</td>
<td>61.9%</td>
<td>Se</td>
</tr>
<tr>
<td>100%</td>
<td>95.8%</td>
<td>0%</td>
<td>100%</td>
<td>100%</td>
<td>Sp</td>
</tr>
<tr>
<td>68.2%</td>
<td>6.7%</td>
<td>0%</td>
<td>35%</td>
<td>47.62%</td>
<td>Se</td>
</tr>
<tr>
<td>100%</td>
<td>95.7%</td>
<td>0%</td>
<td>100%</td>
<td>100%</td>
<td>Sp</td>
</tr>
<tr>
<td>50%</td>
<td>0%</td>
<td>0%</td>
<td>28.6%</td>
<td>28.6%</td>
<td>Se</td>
</tr>
<tr>
<td>100%</td>
<td>100%</td>
<td>0%</td>
<td>100%</td>
<td>100%</td>
<td>Se</td>
</tr>
<tr>
<td>42%</td>
<td>0%</td>
<td>0%</td>
<td>23.8%</td>
<td>23.8%</td>
<td>Se</td>
</tr>
<tr>
<td>100%</td>
<td>100%</td>
<td>0%</td>
<td>100%</td>
<td>100%</td>
<td>Sp</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Baseline-Retest</th>
<th>All Norms</th>
<th>Acute Stringency</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;15th Percentile</td>
<td>&lt;10th Percentile</td>
<td>&lt;5th Percentile</td>
</tr>
<tr>
<td>70 RCI</td>
<td>80 RCI</td>
<td>90 RCI</td>
</tr>
<tr>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>95.7%</td>
<td>95.7%</td>
<td>95.7%</td>
</tr>
<tr>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>50%</td>
<td>30%</td>
<td>0%</td>
</tr>
<tr>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>42%</td>
<td>23.8%</td>
<td>0%</td>
</tr>
<tr>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>
The table below provides a summary of sensitivity and specificity of SCAT3 components individually and combined at the acute phase of injury using a test-retest and normative approach.

<table>
<thead>
<tr>
<th>Stringency</th>
<th>Female Norms</th>
<th>Male Norms</th>
<th>Combined SCAT3</th>
<th>m-BESS</th>
<th>SAC</th>
<th>Symptom Total</th>
<th>Symptom Severity</th>
<th>Utility</th>
<th>Acute</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;1st Percentile</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>66%</td>
<td>100%</td>
<td>66%</td>
</tr>
<tr>
<td>&lt;5th Percentile</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>88.9%</td>
<td>100%</td>
<td>88.9%</td>
</tr>
<tr>
<td>&lt;10th Percentile</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>88.9%</td>
<td>100%</td>
<td>88.9%</td>
</tr>
<tr>
<td>&lt;15th Percentile</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>88.9%</td>
<td>100%</td>
<td>88.9%</td>
</tr>
</tbody>
</table>

The table includes the following components:
- **SCAT3**
- **m-BESS**
- **SAC**
- **Symptom Total**
- **Symptom Severity**
- **Utility**
- **Acute**
### Table A.3: Sensitivity and Specificity of SCAT3 Components Individually and Combined 3 Weeks Post-Injury Using a Test-Retest and Normative Approach

<table>
<thead>
<tr>
<th>Component</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symptom Stringency</td>
<td>96.7%</td>
<td>100%</td>
</tr>
<tr>
<td>Symptom Severity</td>
<td>96.6%</td>
<td>100%</td>
</tr>
<tr>
<td>SAC</td>
<td>96.7%</td>
<td>100%</td>
</tr>
<tr>
<td>m-BESS</td>
<td>91.3%</td>
<td>20%</td>
</tr>
<tr>
<td>Combined SCAT3</td>
<td>61.6%</td>
<td>100%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Component</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Symptoms</td>
<td>8.33%</td>
<td>100%</td>
</tr>
<tr>
<td>Utility</td>
<td>0%</td>
<td>100%</td>
</tr>
<tr>
<td>Post-Acute Baseline-Retest</td>
<td>0.7 RCI</td>
<td>8 RCI</td>
</tr>
</tbody>
</table>

Note: The table includes components of the Sports Concussion Assessment Tool (SCAT3) and their sensitivity and specificity values for various thresholds and approaches. The data is presented in terms of percentiles and RCI values, indicating the performance of each component in detecting and ruling out concussions.
<table>
<thead>
<tr>
<th>Female Norms</th>
<th>Male Norms</th>
<th>Acute</th>
<th>Post-Injury</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symptom Severity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Symptoms</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
</tr>
<tr>
<td>m-BESS</td>
<td>33.3%</td>
<td>100%</td>
<td>33.3%</td>
</tr>
<tr>
<td>SAC</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
</tr>
<tr>
<td>Combined SCAT3</td>
<td>77.8%</td>
<td>100%</td>
<td>77.8%</td>
</tr>
</tbody>
</table>

Table 4.3: Sensitivity and specificity of SCAT3 components individually and combined 3 weeks post-injury using a test-retest and normative approach.
61.1-89.1% at the acute phase of injury, and 52-75% 3 weeks post-injury. The highest sensitivity reached by the m-BESS was 33.3% using the male normative approach 3 weeks post-injury, whereas the SAC reached its highest sensitivity of 15.4% at the acute phase of injury using the baseline retest approach. Importantly, the sensitivity was found to be greatest for the combined SCAT3, reaching a high of 89.1% at the acute phase of injury using the female normative approach. The specificity was high across all tests; reaching 100% for the combined SCAT3 for both the acute and post-acute test phases.

Notably, the sensitivity and specificity of the individual and combined SCAT3 components varied according to the method used for their calculations. Table 4.3 depicts the sensitivity of the independent and combined SCAT3 component scores ranked from one (highest) to four (lowest) across each of the methodologies used at the acute phase of testing.

<table>
<thead>
<tr>
<th>SCAT3 Component</th>
<th>Method (Sensitivity)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total Symptoms</strong></td>
<td>1. Female norms (72.7%)</td>
</tr>
<tr>
<td></td>
<td>2. All norms (61.9%)</td>
</tr>
<tr>
<td></td>
<td>3. Baseline retest (47.4%)</td>
</tr>
<tr>
<td></td>
<td>4. Male norms (44.4%)</td>
</tr>
<tr>
<td><strong>Symptom severity</strong></td>
<td>1. Female norms (60%)</td>
</tr>
<tr>
<td></td>
<td>2. All norms (39.1%)</td>
</tr>
<tr>
<td></td>
<td>3. Baseline retest (36.8%)</td>
</tr>
<tr>
<td></td>
<td>4. Male norms (30%)</td>
</tr>
<tr>
<td><strong>SAC</strong></td>
<td>1. Baseline retest (15.4%)</td>
</tr>
<tr>
<td></td>
<td>2. All norms, Male norms and Female norms all 0%</td>
</tr>
<tr>
<td><strong>m-BESS</strong></td>
<td>1. Baseline retest (15.4%)</td>
</tr>
<tr>
<td></td>
<td>2. Male norms, (12.5%)</td>
</tr>
<tr>
<td></td>
<td>3. All norms (7.14%)</td>
</tr>
<tr>
<td></td>
<td>4. Female norms (0%)</td>
</tr>
</tbody>
</table>
Evidently, the female normative approach led to superior sensitivity values for the total symptom score, the symptom severity score, and the combined SCAT3, however the baseline retest approach yielded highest sensitivity values for the m-BESS and the SAC.

Table 4.5. SCAT3 Sensitivity Ranking by Methodology at 3 Weeks Post-Injury

<table>
<thead>
<tr>
<th>SCAT3 Component</th>
<th>Method (Sensitivity)</th>
</tr>
</thead>
</table>
| Total Symptoms      | 1. Female norms (62.5%)  
                      | 2. All norms (58.3%)  
                      | 3. Male norms (50%)  
                      | 4. Baseline retest (40%) |
| Symptom severity    | 1. Female norms (33.3%)  
                      | 2. Male norms (33.3%)  
                      | 3. All norms (25%)  
                      | 4. Baseline retest (20%) |
| SAC                 | 1. Baseline retest, All norms, Male norms and Female norms all 0%                   |
| m-BESS              | 1. Male norms, (33.3%)  
                      | 2. Baseline retest (20%)  
                      | 3. All norms (14.3%)  
                      | 4. Female norms (0%) |
| Combined SCAT3      | 1. Male norms (77.8%)  
                      | 2. Female norms (75%)  
                      | 3. All norms (73.2%)  
                      | 4. Baseline retest (61.6%) |
Table 4.4 outlines the sensitivity of the independent and combined SCAT3 component scores ranked from one (highest) to four (lowest) across each of the methodologies used at the post-acute phase of injury. By looking at this table, similar results are revealed for the total symptom and symptom severity scores, with the female normative approach once again leading to the highest sensitivity values. The m-BESS sensitivity across each method is also similarly consistent across testing times; with the male normative and a baseline retest approach yielding the highest sensitivity values. Moreover, for the SAC, all methods led to 0% sensitivity when administered 3 weeks post-injury. Lastly, the sensitivity of the combined SCAT3 was superior using the male normative approach, followed by the female normative approach, which is in stark contrast to the findings at the acute phase of injury. However, this low sensitivity in the female normative approach is likely driven by the lack of sensitivity of the m-BESS, leaving only the PCSS scores to contribute to the combined sensitivity.

In comparing the sensitivity of the SCAT3 between testing times, it is evident that sensitivity decreased over time for the total symptom score using all approaches except the male normative approach, for the symptom severity score for all methods, for the SAC across all methods, and for the combined SCAT3 across all methods except the male normative approach. In contrast, the m-BESS became more sensitive when administered 3 weeks post-injury, compared to the acute phase of injury.

Over time, specificity tended to increase across the SCAT3 component scores. Particularly, for the total symptom score, specificity values increased using the baseline retest, all norms, and female norms approach, however decreased using the male norm approach. Symptom severity specificity increased over time using the male and female normative approach, yet decreased using the baseline-retest and all norms approach. For the m-BESS, the specificity increased using the male and all normative approach, while remaining the same (100%) using the baseline retest and female normative approach. Lastly, the SAC specificity increased at 3 weeks post injury compared to the acute phase of injury using the male norms and all norms, however remained the same for the female norms and decreased using the baseline retest approach.

### 4.5.5 Sensitivity and specificity of the ANAM

The reliability values for the ANAM scores based on the healthy control data are presented in table 4.6 for the throughput scores, table 4.7 for the reaction time scores, and table 4.8 for
percent correct items. Tables 4.9 to 4.11 outline the results from the sensitivity and specificity calculations of each ANAM subtest across the main outcome measures of throughput (TP) score, percent correct (PC) and reaction time (RT) for correct responses at the acute phase of injury. Tables 4.12-4.14 outline the results from the sensitivity and specificity calculations of each ANAM subtest across the main outcome measures 3 weeks post-injury. Additionally, Tables 4.15 and 4.16 outline the sensitivity and specificity from combining the TP, PC, and RT for each of the ANAM subtests across the retest and normative approaches at the acute and post-acute phases of injury, respectively.

Table 4.6. Pearson correlation coefficient reliability values for ANAM throughput scores.

<table>
<thead>
<tr>
<th>ANAM Throughput Scores</th>
<th>Baseline: 3-5 Days</th>
<th>Baseline: 3 Weeks</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDD</td>
<td>.591</td>
<td>.172</td>
</tr>
<tr>
<td>CDS</td>
<td>.334</td>
<td>.577</td>
</tr>
<tr>
<td>M2S</td>
<td>.504</td>
<td>.496</td>
</tr>
<tr>
<td>PRO</td>
<td>.351</td>
<td>-.265</td>
</tr>
<tr>
<td>SPD</td>
<td>.688</td>
<td>.352</td>
</tr>
<tr>
<td>SRT</td>
<td>.113</td>
<td>.061</td>
</tr>
<tr>
<td>SR2</td>
<td>.053</td>
<td>-.072</td>
</tr>
<tr>
<td>ST6</td>
<td>.552</td>
<td>-.015</td>
</tr>
</tbody>
</table>

Table 4.7. Pearson correlation coefficient reliability values for ANAM reaction time scores

<table>
<thead>
<tr>
<th>ANAM Reaction Time</th>
<th>Baseline: 3-5 Days</th>
<th>Baseline: 3 Weeks</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDD</td>
<td>.597</td>
<td>.528</td>
</tr>
<tr>
<td>CDS</td>
<td>.221</td>
<td>.431</td>
</tr>
<tr>
<td>M2S</td>
<td>.468</td>
<td>.594</td>
</tr>
<tr>
<td>PRO</td>
<td>.497</td>
<td>-.128</td>
</tr>
<tr>
<td>SPD</td>
<td>.384</td>
<td>.317</td>
</tr>
<tr>
<td>SRT</td>
<td>.127</td>
<td>.085</td>
</tr>
<tr>
<td>SR2</td>
<td>-.008</td>
<td>-.086</td>
</tr>
<tr>
<td>ST6</td>
<td>.597</td>
<td>.04</td>
</tr>
</tbody>
</table>
Table 4.8. Pearson correlation coefficient reliability values for ANAM percent correct items.

<table>
<thead>
<tr>
<th>ANAM Percent Correct</th>
<th>Baseline: 3-5 Days</th>
<th>Baseline: 3 Weeks</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDD</td>
<td>.814</td>
<td>.865</td>
</tr>
<tr>
<td>CDS</td>
<td>.184</td>
<td>.275</td>
</tr>
<tr>
<td>M2S</td>
<td>-.152</td>
<td>-.097</td>
</tr>
<tr>
<td>PRO</td>
<td>.02</td>
<td>-.01</td>
</tr>
<tr>
<td>SPD</td>
<td>.223</td>
<td>-.07</td>
</tr>
<tr>
<td>SRT</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>SR2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>ST6</td>
<td>.351</td>
<td>.124</td>
</tr>
</tbody>
</table>

As large amounts of data are outlined in these tables (4.9-4.16), the main findings will be highlighted here. Once again it can be seen that the .70 confidence interval from the test retest approach and <15th percentile from the normative approaches resulted in the greatest sensitivity values, therefore these results will be discussed more thoroughly.

Firstly, it was found that the sensitivity of each test measure varied across methods of calculation, however the tests of reaction time (SRT, SR2, PRO) remained highly sensitive throughout, whereas the CDS and CDD tests demonstrated the lowest sensitivities across methods. Although the specificity values remained high across most of the ANAM subtest outcome measures, the SRT and SR2 throughput scores and reaction times were found to be considerably low.

From these tables it can also be seen that the throughput score and reaction time for correct responses typically yielded the highest sensitivity values across each of the ANAM subtests, whereas the percent correct contributed little to each of the overall tests’ sensitivity. This outcome measure was especially lacking sensitivity for the SRT and SR2 tasks as no participant received a score of less than 100% at any time point, and as such all athletes were classified as being healthy and the sensitivity remained at 0%.
### Table 4.9. Sensitivity and specificity of ANAM component throughput scores individually and combined 3 to 5 days post-injury using a test-retest and normative approach.

<table>
<thead>
<tr>
<th>Component</th>
<th>Female Norms</th>
<th>Male Norms</th>
<th>Combined Norms</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% 0% 10% 20%</td>
<td>% 0% 10% 20%</td>
<td>% 0% 10% 20% 30%</td>
</tr>
<tr>
<td>ANAM TP</td>
<td>% 0% 10% 20%</td>
<td>% 0% 10% 20%</td>
<td>% 0% 10% 20% 30%</td>
</tr>
<tr>
<td>CDS TP</td>
<td>% 0% 10% 20%</td>
<td>% 0% 10% 20%</td>
<td>% 0% 10% 20% 30%</td>
</tr>
<tr>
<td>SRT TP</td>
<td>% 0% 10% 20%</td>
<td>% 0% 10% 20%</td>
<td>% 0% 10% 20% 30%</td>
</tr>
<tr>
<td>SPD TP</td>
<td>% 0% 10% 20%</td>
<td>% 0% 10% 20%</td>
<td>% 0% 10% 20% 30%</td>
</tr>
<tr>
<td>M2S TP</td>
<td>% 0% 10% 20%</td>
<td>% 0% 10% 20%</td>
<td>% 0% 10% 20% 30%</td>
</tr>
<tr>
<td>ST6 TP</td>
<td>% 0% 10% 20%</td>
<td>% 0% 10% 20%</td>
<td>% 0% 10% 20% 30%</td>
</tr>
<tr>
<td></td>
<td>% 0% 10% 20%</td>
<td>% 0% 10% 20%</td>
<td>% 0% 10% 20% 30%</td>
</tr>
</tbody>
</table>

The table above shows the sensitivity and specificity of ANAM component throughput scores individually and combined 3 to 5 days post-injury using a test-retest and normative approach.
Table 4.10. Cont. Sensitivity and specificity of ANAM component percent correct scores individually and combined 3 to 5 days post-injury using a test-retest and normative approach.

<table>
<thead>
<tr>
<th>Component</th>
<th>Male Norms</th>
<th>Female Norms</th>
<th>Acute</th>
<th>Baseline-Retest</th>
<th>Utility</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANAM RT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST6 RT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRT RT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPD RT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRO RT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M2S RT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDS RT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDD RT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Utility</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specificity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strenuity</th>
<th>&lt;5th Percentile</th>
<th>&lt;10th Percentile</th>
<th>&lt;20th Percentile</th>
<th>&lt;30th Percentile</th>
<th>&lt;40th Percentile</th>
<th>&lt;50th Percentile</th>
<th>&lt;60th Percentile</th>
<th>&lt;70th Percentile</th>
<th>&lt;80th Percentile</th>
<th>&lt;90th Percentile</th>
<th>&lt;100th Percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utility</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specificity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strenuity</th>
<th>&lt;5th Percentile</th>
<th>&lt;10th Percentile</th>
<th>&lt;20th Percentile</th>
<th>&lt;30th Percentile</th>
<th>&lt;40th Percentile</th>
<th>&lt;50th Percentile</th>
<th>&lt;60th Percentile</th>
<th>&lt;70th Percentile</th>
<th>&lt;80th Percentile</th>
<th>&lt;90th Percentile</th>
<th>&lt;100th Percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utility</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specificity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.11. Cont. Sensitivity and specificity of ANAM component reaction time scores individually and combined 3 to 5 days post-injury using a test-retest and normative approach.

<table>
<thead>
<tr>
<th>Component</th>
<th>Male Norms</th>
<th>Female Norms</th>
<th>Acute</th>
<th>Baseline-Retest</th>
<th>Utility</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANAM RT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST6 RT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRT RT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPD RT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRO RT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M2S RT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDS RT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDD RT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Utility</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specificity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strenuity</th>
<th>&lt;5th Percentile</th>
<th>&lt;10th Percentile</th>
<th>&lt;20th Percentile</th>
<th>&lt;30th Percentile</th>
<th>&lt;40th Percentile</th>
<th>&lt;50th Percentile</th>
<th>&lt;60th Percentile</th>
<th>&lt;70th Percentile</th>
<th>&lt;80th Percentile</th>
<th>&lt;90th Percentile</th>
<th>&lt;100th Percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utility</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specificity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strenuity</th>
<th>&lt;5th Percentile</th>
<th>&lt;10th Percentile</th>
<th>&lt;20th Percentile</th>
<th>&lt;30th Percentile</th>
<th>&lt;40th Percentile</th>
<th>&lt;50th Percentile</th>
<th>&lt;60th Percentile</th>
<th>&lt;70th Percentile</th>
<th>&lt;80th Percentile</th>
<th>&lt;90th Percentile</th>
<th>&lt;100th Percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utility</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specificity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Table 1</td>
<td>Sensitivity and specificity of ANAM component throughput scores individually and combined 3 weeks post-injury using a test-retest and normative approach.</td>
<td>100.0%</td>
<td>12.5%</td>
<td>93.8%</td>
<td>62.5%</td>
<td>100.0%</td>
<td>0.0%</td>
<td>100.0%</td>
<td>81.3%</td>
<td>20.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Combined TP</td>
<td>Sp</td>
<td>Se</td>
<td>Sp</td>
<td>Se</td>
<td>Sp</td>
<td>Se</td>
<td>Sp</td>
<td>Se</td>
<td>Sp</td>
<td>Se</td>
<td>Sp</td>
</tr>
<tr>
<td>100.0%</td>
<td>0.0%</td>
<td>93.3%</td>
<td>0.0%</td>
<td>37.5%</td>
<td>62.5%</td>
<td>100.0%</td>
<td>0.0%</td>
<td>93.3%</td>
<td>0.0%</td>
<td>86.7%</td>
<td>100.0%</td>
</tr>
<tr>
<td>SR2 TP</td>
<td>Sp</td>
<td>Se</td>
<td>Sp</td>
<td>Se</td>
<td>Sp</td>
<td>Se</td>
<td>Sp</td>
<td>Se</td>
<td>Sp</td>
<td>Se</td>
<td>Sp</td>
</tr>
<tr>
<td>100.0%</td>
<td>0.0%</td>
<td>100.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>100.0%</td>
<td>0.0%</td>
<td>93.3%</td>
<td>0.0%</td>
<td>93.3%</td>
<td>0.0%</td>
<td>93.3%</td>
</tr>
<tr>
<td>SPD TP</td>
<td>Sp</td>
<td>Se</td>
<td>Sp</td>
<td>Se</td>
<td>Sp</td>
<td>Se</td>
<td>Sp</td>
<td>Se</td>
<td>Sp</td>
<td>Se</td>
<td>Sp</td>
</tr>
<tr>
<td>0.0%</td>
<td>100.0%</td>
<td>0.0%</td>
<td>25.0%</td>
<td>75.0%</td>
<td>37.5%</td>
<td>100.0%</td>
<td>0.0%</td>
<td>80.0%</td>
<td>0.0%</td>
<td>66.7%</td>
<td>20.0%</td>
</tr>
<tr>
<td>PRO TP</td>
<td>Sp</td>
<td>Se</td>
<td>Sp</td>
<td>Se</td>
<td>Sp</td>
<td>Se</td>
<td>Sp</td>
<td>Se</td>
<td>Sp</td>
<td>Se</td>
<td>Sp</td>
</tr>
<tr>
<td>100.0%</td>
<td>0.0%</td>
<td>100.0%</td>
<td>0.0%</td>
<td>100.0%</td>
<td>25.0%</td>
<td>93.8%</td>
<td>25.0%</td>
<td>100.0%</td>
<td>0.0%</td>
<td>86.7%</td>
<td>0.0%</td>
</tr>
<tr>
<td>M2S TP</td>
<td>Sp</td>
<td>Se</td>
<td>Sp</td>
<td>Se</td>
<td>Sp</td>
<td>Se</td>
<td>Sp</td>
<td>Se</td>
<td>Sp</td>
<td>Se</td>
<td>Sp</td>
</tr>
<tr>
<td>100.0%</td>
<td>0.0%</td>
<td>100.0%</td>
<td>12.5%</td>
<td>100.0%</td>
<td>0.0%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>0.0%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>CDD TP</td>
<td>Sp</td>
<td>Se</td>
<td>Sp</td>
<td>Se</td>
<td>Sp</td>
<td>Se</td>
<td>Sp</td>
<td>Se</td>
<td>Sp</td>
<td>Se</td>
<td>Sp</td>
</tr>
<tr>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>12.5%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

Table 4.12. Sensitivity and specificity of ANAM component throughput scores individually and combined 3 weeks post-injury using a test-retest and normative approach.
<p>| Male Norms | Female Norms | Utility | | Stringency | | Female | Male | Utility | Stringency | Female | Male | Utility | Stringency |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PC | PC | PC | PC | PC | PC | PC | PC | PC | PC | PC | PC | PC | PC | PC |
| T| T| T| T| T| T| T| T| T| T| T| T| T| T| T|
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th>0%</th>
<th>10%</th>
<th>20%</th>
<th>30%</th>
<th>40%</th>
<th>50%</th>
<th>60%</th>
<th>70%</th>
<th>80%</th>
<th>90%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.15: Sensitivity and specificity of combined throughput, percent correct, and reaction time scores for each ANAM subtest at 3 weeks post-injury.

Table 4.16: Sensitivity and specificity of combined throughput, percent correct, and reaction time scores for each ANAM subtest at acute phase of injury.

Table 4.17: Combined utility (TP + PC + RT) for baseline and retest for each ANAM subtest.
<table>
<thead>
<tr>
<th>Subtest</th>
<th>Combined Utility (TP + PC + RT) Male Norms</th>
<th>Combined Utility (TP + PC + RT) Female Norms</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt;10th Percentile</td>
<td>&lt;5th Percentile</td>
</tr>
<tr>
<td></td>
<td>&lt;1st Percentile</td>
<td>&lt;15th Percentile</td>
</tr>
<tr>
<td>CDD</td>
<td>33.3%</td>
<td>99.8%</td>
</tr>
<tr>
<td>SRT</td>
<td>33.3%</td>
<td>99.9%</td>
</tr>
<tr>
<td>SR2</td>
<td>33.3%</td>
<td>0.0%</td>
</tr>
<tr>
<td>ST6</td>
<td>33.3%</td>
<td>100.0%</td>
</tr>
<tr>
<td>PRO</td>
<td>33.3%</td>
<td>0.0%</td>
</tr>
<tr>
<td>SPD</td>
<td>33.3%</td>
<td>100.0%</td>
</tr>
<tr>
<td>M2S</td>
<td>33.3%</td>
<td>0.0%</td>
</tr>
<tr>
<td>CDS</td>
<td>33.3%</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

Table 4.16: Sensitivity and specificity of combined throughput, percent correct, and reaction time scores for each ANAM subtest at 3 weeks post-injury.
Moreover, similar to what was observed in the SCAT3 results, the sensitivity of the ANAM subtests tended to decrease over time using the normative approaches. However, it was revealed that using the baseline retest approach led to five of the eight ANAM subtests increasing in sensitivity between the acute and post-acute phases of testing. The specificity calculated across all methodologies resulted in overall higher values over time; similar to what was observed in the SCAT3 results.

In comparing between the different approaches used to calculate the sensitivity and specificity of each of the ANAM subtests, a mix of results was observed. The sensitivity ranking between methods of calculation for each of the ANAM outcome measures at the acute phase of injury and at 3 weeks post-injury are presented in Table 4.17 and Table 4.18, respectively. These tables demonstrate that the normative approach typically yielded higher sensitivity values compared to the baseline retest approach, however the superior sensitivity values in the normative approach were largely contributed to by the SRT and SR2 tasks, which also yielded considerably low sensitivity values that must be taken into consideration when interpreting these results. Nevertheless, only slight differences between methods were revealed.

Table 4.17. ANAM Sensitivity Ranking by Methodology at Acute Phase of Injury

<table>
<thead>
<tr>
<th>ANAM Outcome Measure</th>
<th>Method (Sensitivity)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Throughput score</td>
<td>1. Female norms (100%)</td>
</tr>
<tr>
<td></td>
<td>Male norms (100%)</td>
</tr>
<tr>
<td></td>
<td>2. Baseline retest (81.4%)</td>
</tr>
<tr>
<td></td>
<td>3. All norms (80.72%)</td>
</tr>
<tr>
<td>Percent Correct</td>
<td>1. Male norms (76.1%)</td>
</tr>
<tr>
<td></td>
<td>2. Female norms (57.5%)</td>
</tr>
<tr>
<td></td>
<td>3. All norms (56.5%)</td>
</tr>
<tr>
<td></td>
<td>4. Baseline retest (53.9%)</td>
</tr>
<tr>
<td>Reaction Time</td>
<td>1. Female norms (100%)</td>
</tr>
<tr>
<td></td>
<td>1. Male norms (100%)</td>
</tr>
<tr>
<td></td>
<td>2. All norms (99.8%)</td>
</tr>
<tr>
<td></td>
<td>3. Baseline-retest (85.9%)</td>
</tr>
</tbody>
</table>
Table 4.18. ANAM Sensitivity Ranking by Methodology 3 Weeks Post-Injury

<table>
<thead>
<tr>
<th>ANAM Outcome Measure</th>
<th>Method (Sensitivity)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Throughput score</td>
<td>1. Male norms (100%)</td>
</tr>
<tr>
<td></td>
<td>2. Female norms (99.7%)</td>
</tr>
<tr>
<td></td>
<td>3. Baseline retest (81.4%)</td>
</tr>
<tr>
<td></td>
<td>4. All norms (80.7%)</td>
</tr>
<tr>
<td>Percent Correct</td>
<td>1. Male norms (69.9%)</td>
</tr>
<tr>
<td></td>
<td>2. Female norms (59.8%)</td>
</tr>
<tr>
<td></td>
<td>3. All norms (56.5%)</td>
</tr>
<tr>
<td></td>
<td>4. Baseline retest (42.4%)</td>
</tr>
<tr>
<td>Reaction Time</td>
<td>1. Female norms (99%)</td>
</tr>
<tr>
<td></td>
<td>2. All norms (93.4%)</td>
</tr>
<tr>
<td></td>
<td>3. Male norms (83.3%)</td>
</tr>
<tr>
<td></td>
<td>4. Baseline-retest (78.2%)</td>
</tr>
</tbody>
</table>

Lastly, in comparing between the overall sensitivity and specificity of the ANAM and SCAT3, it is evident that the ANAM is a more sensitive and specific measure when all tests are combined for both the acute and post-acute time points post-injury. Moreover, in comparing the cognitive measure included in the SCAT3, the SAC, a clear difference in the sensitivity is observed between the ANAM cognitive measures, with the ANAM being much more sensitive across all test times compared to the SAC.

4.5.6 Optimal SCAT3 cut-points

The results from the ROC curve analyses across the retest and normative methods are presented in table 4.15 for the acute of injury and table 4.16 for the post-acute phase of injury. Analysis of the coordinate points of the ROC curve revealed that the greatest sensitivity of each test measure resulted from either a 1-point increase or decrease in test score, depending on the nature of the test- i.e. an increase in score is associated with worsening (PCSS) or improvement (SAC, m-BESS).
Similar to the RCI and normative calculations presented earlier, the total symptom and symptom severity scores yielded the highest sensitivity values with this change score, resulting in a significant AUC at the acute and post-acute phase of injury across the female norms (acute: TS: p = .002, SS: p = .003; post acute: TS: p = .016, SS: p = .027), male norms, (acute: TS: p = .002, SS: p = .011; post acute: TS: p = .011, SS: p = .028), and all norms approaches (acute: TS: p = .000, SS: p = .000; post acute: TS: p = .000, SS: p = .002), with the baseline-retest only resulting in a significant AUC in total symptom and symptom severity scores at the acute time post-injury (TS: p = .000, SS: p = .013). Across all of the methodologies, it was found that the female normative approach created the most sensitive values for the 1-point change in score. Within the normative approach, comparison between the median normative score to the athlete’s score led to greater sensitivity as opposed to using the means, with this being anticipated due to the non-normal distribution of the data. Lastly, this change score resulted in decreased sensitivity and increased specificity of the total symptom and symptom severity scores over time, with the opposite being found for the SAC and m-BESS, such that sensitivity tended to stay the same or increase over time while specificity decreased. This result is equivalent to what was found previously with the RCI and normative approach.

4.5.7 SCAT3 ROC Comparisons

In order to statistically quantify which approach- either baseline retest or normative- and which component on the SCAT3 yielded the most sensitive test results, the area under the ROC curves were compared following Delong & Carolina’s (2016) methodology.

There were no differences observed in the AUCs between the normative mean or median approaches, therefore the results presented illustrate one value for each of the normative methods rather than two.

In comparing between the baseline retest and normative approaches, only limited significant differences were found. Firstly, for the total symptom score, no significant differences were revealed at the acute phase of injury between the normative and retest approaches; however the AUCs were greater in the full normative dataset and for the male normative dataset compared to retest, whereas the AUCs were greater in the baseline retest approach compared to the female norms. At the post-acute phase of injury, there were no significant differences between
methodologies, however all of the normative approaches yielded considerably greater AUCs compared to the baseline retest approach.

Similarly, for the symptom severity score, there were no significant differences observed between the normative or baseline retest approaches at either the acute or post-acute phase of injury. However, once again the normative approaches yielded considerably greater AUCs than the retest method, with the complete normative dataset approaching significance, \( z = 1.903, p = .0571 \), Baseline: AUC= .661, SE= .117; All norms: AUC= .877, SE= .0571.

Additionally, there were no significant differences between methods on the SAC at either the acute or post-acute phase of injury. Acutely, the baseline retest approach yielded slightly greater AUCs than the normative approaches, whereas post-acutely the normative method led to greater AUCs compared to retest.

Finally, no significant differences in ROCs were found between the normative or retest approaches on the m-BESS at the acute phase of injury, however the AUCs were considerably greater under the normative approaches compared to the retest approach. Post-acutely, it was revealed that there was a significantly greater AUC in the m-BESS baseline retest method (AUC= .783, SE= .225) compared to the female normative approach (AUC= .60, SE= .217), \( z = 2.668, p = .0076 \). The difference between ROC curves is presented graphically in Figure 4.8. Although the AUCs from the other normative datasets did not differ significantly from the baseline retest approach, the AUCs were similarly greater in the retest compared to the normed method.
In comparing between the different SCAT3 components, a few significant findings were revealed. Firstly, at the acute phase of injury, the analysis within the male normative approach revealed that the AUC in the SAC (AUC = .5, SE = .153) was significantly lower than both the symptom severity (AUC = .833, SE = .106, z = 2.19, p = .0285) and total symptom score (AUC = .903, SE = .0758, z = 2.872, p = .0041). The difference in ROC curves is presented graphically in Figure 4.9.
Figure 4.9. Significantly lower SAC AUC compared to total symptom and symptom severity score at acute phase of injury in male normative data.

Moreover, in the entire normative dataset, it was found that the AUC of the total symptom score (AUC= .819, SE= .0679) was significantly greater than the symptom severity score (AUC= .765, SE= .0762, z= 2.224, p= .0261) the SAC (AUC= .520, SE= .099, z= 2.271, p= .0231), as well as the m-BESS (AUC= .567, SE= .00963, z= 2.101, p= .0356) at the acute phase of injury. The difference between the ROC curves is presented graphically in Figure 4.10. No other significant differences were found between the SCAT3 components at the acute phase of injury within the baseline retest method or the female normative approach. However, in the retest approach it was found that the total symptom score yield the greatest AUC, followed by the symptom severity score, the SAC, and finally the m-BESS. Additionally, across all of the normative approaches it was found that the total symptom score produced the greatest AUC, followed by the symptom severity score, the m-BESS, and the SAC.
Figure 4.10. Significantly greater total symptom AUC compared to symptom severity, SAC, and m-BESS at acute phase of injury in entire normative dataset.

At the post-acute phase of injury, no significant differences were found between the SCAT3 components across either the normative or retest approaches. For the retest approach the SAC was found to have the greatest AUC, followed by the symptom severity score, the total symptom score, and the m-BESS. For the female norms and the complete normative dataset, the SAC had the greatest AUC, followed by the total symptom score, symptom severity score, and then the m-BESS. Lastly, for the male normative approach, the total symptom score yielded the greatest AUC, followed by the symptom severity score, the SAC, and the m-BESS.

4.6 Discussion

The current study sought to extend the prior work of Putukian and colleagues (2015) by assessing the clinical utility of the SCAT3 in identifying concussion in varsity athletes across
The results from this study revealed that the most sensitive measures included in the SCAT3 were the total symptom and symptom severity scores comprised in the PCSS. The sensitivity of the total symptom score ranged across methodologies of 44.4%- 72.7% at the acute phase of injury and 40-62.5% at 3 weeks post-injury. The sensitivity of the symptom severity scale ranged between 30-60% at the acute phase of injury when baseline data was or was not made available, and between 20-33.3% at 3 weeks post-injury. The combined sensitivity of the PCSS total symptom and symptom severity scores ranged from 61.1-89.1% at the acute phase of injury, and 52-75% 3 weeks post-injury.

As the sensitivity of the exact iteration included in the SCAT3 has never been evaluated before, these results were found to corroborate previous research conducted with other symptom scales finding moderate to high sensitivity at the acute phase of injury and low to moderate sensitivity as time post-injury extends. Specifically, McCrea and colleagues (2005) demonstrated that a 17-item GSC was 89% sensitive and 100% specific in identifying concussion at the time of injury, dropping to 74% post-game and to 4% seven days post-injury.

This study is also the first of its kind to statistically assess for significant differences across the SCAT3 components over time using a baseline and normative approach. When the entire normative dataset was used to compare against post-concussion scores within 3 to 5 days, the total symptom score yielded a significantly greater area under the ROC curve compared to the symptom severity, SAC, and m-BESS measures. This finding further substantiates the utility of the total symptom score in identifying concussion in the acute phase of injury.

Moreover, the significant finding of elevated symptom scores in concussed athletes at the acute phase of injury compared to baseline and in comparison to healthy controls is a result that is well established in the literature, however has now been replicated using the PCSS in the SCAT3 specifically (Mayfield, Bay, & McLeod, 2013; Resch et al., 2016).
SAC
The results from this study also highlight the limited utility contributed by the SAC in identifying concussion if more than 72 hours post-concussion, finding that sensitivity ranged from 0-15.4% at the acute phase of injury and 0% across all methods at 3 weeks post-injury. Additionally, comparison of the SCAT3 component ROC curves revealed that in the male normative dataset, the SAC AUC was significantly lower than the total symptom and symptom severity scores in the acute phase of injury, further highlighting the limited utility of this tool in comparison to other concussion assessment measures. This finding supports previous research outlining the low sensitivity of the SAC if used external to the sideline, with the SAC identifying impairments in 31% of concussed athletes at 24 hours post-injury, 23% at 48 hours, and 9% if tested at one week post-injury (McCrea et al., 2005). This study is the first of its kind to assess the utility of the SAC in identifying concussed athletes at 3 weeks post-injury, however follows the trend in the literature of being less sensitive as the time post-injury extends.

m-BESS
The utility of the m-BESS in identifying concussion over time has never been previously evaluated, however the results of this study corroborate previous research on the validity of the original BESS, finding that it lacks sensitivity after the third day post-concussion (Guskiewicz, Ross, & Marshall, 2001; Riemann & Guskiewicz, 2000). This study found that the m-BESS has low sensitivity, ranging between 0-15.4% within the first 3 to 5 days following concussion and 0-33.3% at 3 weeks post-injury.

Interestingly, the sensitivity of the m-BESS increased over time using the male normative approach and the baseline-retest method, however only modestly. As previous research has found that the original BESS is useful in detecting gross balance deficits within 24 hours post-injury (Guskiewicz, 2001; Riemann, Guskiewicz, & Shields, 1999), it could be reasoned that sensitivity increased due to pronounced postural instabilities in the male athletes who remained concussed at 3 weeks post-injury. Additionally, the total number of false negatives (identifying a concussed athlete as being healthy) that could be made naturally decreased over time as the number of healthy athletes increased, with concussed athletes becoming medically cleared, augmenting the potential for increased sensitivity (Please reference sensitivity calculation for understanding: Sensitivity= True Positives/ (True Positives + False Negatives). Therefore, the
sensitivity of the m-BESS appeared to increase over time due to this lesser probability of making a false negative error, in combination with the m-BESS capturing the athletes’ gross balance impairments.

**Combined SCAT3**

In evaluating the clinical utility of the SCAT3, it is critical to assess the combined sensitivity and specificity of the measure. The results from this research demonstrate that the SCAT3 as a multidimensional concussion assessment tool is clinically useful across the concussion sequelae. Combined, the sensitivity of the SCAT3 ranged between 66-89.1% within the first 3 to 5 days post-injury and 61.6-77.8% at 3 weeks post-injury, with the specificity remaining at 100% across the recovery trajectory. This finding substantiates previous research demonstrating heightened sensitivity when multiple concussion screening tools are combined (Broglio & Puetz, 2008; Giza et al., 2013), and lends support to the recommended use of a multifaceted approach to identifying concussion.

**ANAM**

This study is also the first of its kind to address the utility of the SCAT3 in comparison to a commonly used computerized neurocognitive assessment battery, the ANAM.

The findings from this research revealed that ANAM as a combined measure is more sensitive to detecting concussion across the recovery trajectory compared to the SCAT3. In directly comparing the cognitive measures, the SAC was found to be considerably less sensitive than the ANAM. As the measures found to be most sensitive to identifying concussion in the ANAM included tests of reaction time, it is reasonable that the SAC displayed low sensitivity given that reaction time was not a component assessed for in the cognitive test. This finding is consistent with previous research demonstrating the utility of the ANAM and the SRT particularly, in detecting concussion within 72 hours post-injury (Coldren, Russell, Parish, Dretsch, & Kelly, 2012). However the current results also contrast previous research demonstrating poor sensitivity of the ANAM using a baseline-retest approach, with no independent measure reaching a sensitivity greater than 14.7% (Register-Mihalik et al., 2013). Although limited sensitivity of the ANAM was found in this previous research, the SRT once again demonstrated the greatest sensitivity, signifying the importance of measuring reaction time.
As the SCAT3 was developed for use on the sideline, the measures comprising the SCAT3 only included those that are easy to administer. Accordingly, measuring reaction time may not be feasible on the sideline and so a more comprehensive measure like the ANAM is critical to administer across the concussion sequelae to help inform clinical decision-making.

Sensitivity of SCAT3 With and Without Baseline Scores
The findings discovered in comparing between the sensitivity of the SCAT3 with and without baseline data are salient. Specifically, the results demonstrate that the sensitivity of the PCSS is heightened using a normative approach compared to using baseline data. The opposite was found for the SAC and m-BESS, with the use of baseline data resulting in improved sensitivity compared to a normative approach, however still demonstrating overall poor sensitivity. These findings address the nature of the assessments being conducted; whereas symptoms typically fluctuate depending on the state of the athlete, gross cognitive and balance functioning are less variable and are more consistent within an individual over time.

As such, the utility of the PCSS in the SCAT3 may be beneficially impacted from using a normative approach that standardizes and moderates the instability of scores to compare post-concussion scores against. As this result was not found using the male normative approach, it could be conceived that males are less variable in symptom reporting over multiple testing times and therefore may benefit from the use of baseline data to compare post-concussion scores directly against. Indeed, by assessing the variability in symptom total score and severity score, it was found that there is less variability in symptom reporting in male athletes compared to females.

Moreover, the sensitivity and specificity analyses corroborate earlier findings demonstrating fewer significant differences between groups in the male athletes compared to the between group findings in females, and more significant differences within the male concussed athletes across time compared to the within group analyses in the female athletes. This suggests that it is important to compare male athletes to their own test scores post-concussion, however females might benefit from a normative approach.
Importantly, although the results presented demonstrate sex-based differences in the sensitivity of the SCAT3 components independently and as a combined measure, the results should be interpreted with caution due to the reduced sample size from stratifying males and females. Indeed, a recent meta-analysis evaluating the differences in symptom reporting between males and females at baseline and post-concussion revealed divergent symptom reporting, however clinically insignificant differences, substantiating such precautions (Brown, Elsass, Miller, Reed, & Reneker, 2015).

Lastly, although differences in the utility of the SCAT3 components independently and combined were observed between normative and baseline methods of calculation, there were limited statistically significant differences found when comparing the areas under the ROC curves. This lack of a statistically significant difference between the baseline and normative ROC curves provides insight into the questionable utility of baseline SCAT3 data. Specifically, as no significant differences were found when baseline data was or was not used, the implementation of baseline testing with the SCAT3 cannot be recommended.

Limitations
The current study is not without limitations. Firstly, although most analyses consisted of the full sample of 45 participants, there were some instances where data were missing for participants. For instance, in the concussed sample, two athletes did not have baseline testing completed and one athlete was not assessed at three weeks post-injury due to attrition. Moreover, within the concussed sample, five athletes only completed the PCSS portion of the SCAT3 upon retest, without completing the SAC or m-BESS. However, as data from the full sample of participants was used in most analyses the results should continue to be considered salient.

Another limitation to this study is the inherent potential for deception at baseline on the subjective symptom reports that could alter post-concussion interpretation. Specifically, augmenting symptoms at baseline in the hopes of returning to play faster after sustaining a concussion can reduce the chance of finding any significant change in scores. Indeed, this was revealed in the healthy control sample, where the reported number and severity of symptoms decreased between baseline and retest. The use of reliable change indices taken from the control
sample in the current study may be a solution to this inherent deception by moderating its effect. Additionally, recent research has demonstrated that repeat assessment after initial invalid baseline neurocognitive performance can improve validity of test scores (Bruce et al., 2016; Schatz et al., 2014). Accordingly, future studies may address this issue by administering two different baseline assessments if feasible.

Another limitation of the study was that the control group of athletes was not entirely considered “healthy” in the sense that 5 of the 22 athletes reported a previous history of mental health, compared to the concussed athletes, although this did not reach significance. Although none of the concussed athletes had a history of mental health issues, 5 of the 23 concussed athletes reported a familial history of mental health, with 8 of the 22 control athletes reporting having a family member previously diagnosed with a mental health disorder. Initially, control athletes were matched to the concussed athletes on the characteristics of sport, sex, age, and history of concussion, however after recognizing this trend of mental health history, only athletes without a previous diagnosis of a mental health disorder were recruited to participate as controls. As the results from the previous chapter revealed that a previous diagnosis of a mental health disorder significantly moderated baseline SCAT3 symptom scores, this could have potentially moderated the power associated with clinical post-concussion test interpretation. However, when comparing between the control and concussed athletes at baseline, no significant differences were found on any of the SCAT3 component scores, including the total symptom and symptom severity scores. Additionally, when re-analyzing the data after removing the 5 control participants with a history of mental health, only minor changes were observed, with the sensitivity of the symptom severity score increasing in the baseline-retest approach under the .70 confidence interval, and no changes being observed on any of the other SCAT3 component scores. Moreover, re-analysis of the ROC curve coordinate points and AUC comparisons revealed no differences in the statistical findings.

Moreover, efforts were made to recruit control participants close in proximity to the time of the concussed athletes in order to regulate academic and varsity pressures, this was not always possible. As such, in some instances differences in test scores might reflect stressors or fatigue related to the time of testing in the academic year or athletic season.
Lastly, it is important to highlight the potential limitation of using statistical conclusions to inform clinical decisions. Clinical utility refers to whether the research findings are significant enough to alter practice. In this sense, statistics alone cannot provide an answer to this question. Instead the clinician must consider how beneficial it is to alter their clinical approach, as well as the feasibility and cost associated with altering one’s practice. While although statistical significance does not equate to clinical utility, the results from this research may still be used to assist in making clinically relevant decisions.

4.7 Conclusions

In summary, the findings from this study support the clinical use of the SCAT3 as a brief multifaceted assessment tool to identify sport-related concussion across the concussion recovery continuum. The use of baseline test scores appears useful for identifying reliable changes post-concussion on the brief cognitive and balance measures included in the SCAT3, however due to the inherent instability of symptom reporting within an athlete over time, the use of normative data may prove beneficial when interpreting post-concussion scores on the PCSS. The lack of a significant difference between the two approaches substantiates expert guidelines in not endorsing baseline testing as mandatory (McCrory et al., 2013).

The total symptom and symptom severity scores yielded the greatest sensitivity and specificity in identifying sport-related concussion compared to the SAC and m-BESS, suggesting that symptom assessment is most pivotal to understanding concussion across the recovery continuum. The results from this research may be used to assist in making clinically relevant decisions regarding the SCAT3. Specifically, it is recommended that the SCAT3 be used in its entirety at the acute phase of injury, with the PCSS being administered independently beyond 3 to 5 days following the concussive injury.

When comparing against the ANAM, it was revealed that the reaction time tests within the computerized neurocognitive battery were more sensitive to detecting concussion compared the SCAT3. This demonstrates the need for a multifaceted and comprehensive approach to identifying sport-related concussion across the recovery trajectory.
Chapter 5
General Discussion

The purpose of the current set of studies was so to update previous research and contribute meaningful information pertaining to the psychometric properties and clinical utility of the SCAT3. Presently, there are no stand-alone objective measures available to identify concussion. The recently developed SCAT3 provides a strong framework for evaluating concussion by utilizing a multifaceted approach, assessing commonly reported signs and symptoms, postural ability, and cognitive performance. Although the SCAT3 is widely acknowledged as the recommended measure used to evaluate sport-related concussion on the sideline and in concussion clinics, there has been limited research validating the clinical utility of this measure as a whole over time.

The objectives of this research were twofold. Primarily, this research sought to evaluate the clinical utility of the SCAT3 by defining the sensitivity and specificity of the measures’ subtests independently and in combination across time using both a normative and baseline-retest approach, as well as in comparison to a commonly used computerized neurocognitive testing battery. Next, this research sought to establish a normative database for the SCAT3 in a Canadian population and to identify potential moderators of SCAT3 baseline score that may affect post-concussion test interpretation.

5.1 Normative SCAT3 values and modifying factors in varsity athletes playing a high-risk sport

Chapter 3 established demographically normative data for the primary SCAT3 component scores in a Canadian varsity athletic sample across the factors of: sex, years of education/age, history of migraine headaches, history of a diagnosed psychiatric disorder personally or within the family, previous diagnosis with LD/ADHD, and history of concussion. The results of this study demonstrated that the components of the SCAT3 are largely affected by an athlete’s demographic characteristics and health history.
Specifically, the total symptom and symptom severity scores comprised in the PCSS of the SCAT3 were found to be affected by sex, individual and familial mental health history, age and years of education, and concussion history. It was found that males compared to females, athletes with a previous personal or familial history of mental health issues compared to those without, younger athletes compared to older athletes, and athletes with 1-2 previous concussions compared to those without reported significantly more symptoms of greater severity on the PCSS. Amongst these factors, it was found that mental health history and years of education accounted for most of the variance in the observed heightened symptoms at baseline, elucidating previous inconsistencies on the impact of sex on baseline symptom reporting. These findings are salient as recent research has revealed that pre-injury SCAT3 symptom severity scores predict acute post-concussion symptom severity, providing insight into clinical recovery time (Nelson et al., 2016). Athletes observed to have heightened symptom severity scores at baseline could therefore be followed more thoroughly upon sustaining a concussion to facilitate management and optimize recovery.

Furthermore, the SAC score was moderated by sex, history of LD/ADHD, and an interaction between sex and history of concussions, such that males performed significantly worse than females, males with 1-2 previous concussions performed worse than females without any previous concussion, and athletes with a history of LD/ADHD performed significantly worse compared to athletes with no previous diagnosis. Amongst these variables, sex was found to account for most of the variance in baseline SAC scores. This finding is consistent with previous research demonstrating greater performance on the SAC in females compared to males in a collegiate sample (Zimmer et al., 2014).

Lastly, this study revealed that baseline m-BESS performance was not influenced by sex, age, history of concussion, history of migraine headaches, LD/ADHD, or personal or familial mental health history.

Establishing a normative database of SCAT3 values in a Canadian varsity athletic population and defining a range of factors that influence baseline SCAT3 scores are paramount to understanding SCAT3 post-concussion test interpretation. The developers of the SCAT3 recommend baseline
test implementation, however if baseline data is lacking, this diverse sample of 524 athletes may be used as a reference to guide clinical decisions post-concussion.

5.2 Clinical utility of the SCAT3

5.2.1 SCAT3

Chapter 4 utilized the normative database established in the previous chapter to evaluate the clinical utility of the SCAT3 in identifying concussion across time in both the presence and absence of baseline test scores and in comparison to the ANAM neurocognitive testing battery.

The key findings of this study demonstrated that the PCSS total symptom and symptom severity scores were the most sensitive and specific in identifying concussion across the recovery trajectory, followed by the m-BESS, and finally the SAC. The sensitivity of the SCAT3 was moderate to high across the concussion sequelae, ranging between 66-89.1% within the first 3 to 5 days post-injury and 61.6- 77.8% at 3 weeks post-injury, with the specificity remaining at 100% over the testing times and across the baseline-retest and normative approaches.

Although the SCAT3 was initially intended for use on the sideline, this finding lends support to the use of the SCAT3 across time within sports management practices to facilitate in clinical decision-making. It is often suggested that the sensitivity of a diagnostic tool be enhanced at the acute phase post-concussion in order to reduce the risk of putting an athlete back into play when they are potentially still concussed, leading to the exacerbation of symptom, while it is recommended that specificity be enhanced as time post-injury extends in order to reduce the risk of withholding an athlete from play when they are truly healthy. Such recommendations are in line with the findings of this study, with the sensitivity of the SCAT3 being heightened within the first 3 to 5 days post-concussion and reducing over time, with specificity increasing across the recovery trajectory. As such, the SCAT3 appears suitable for distinguishing between healthy and concussed athletes across the concussion sequelae.

It was also found that a 1-point change in the SCAT3 component scores yielded a sensitivity of 60-100% for the total symptom score, 50-81.8% for the symptom severity score, 14.3-42.9% for the SAC, and 14.3-43.7% across the baseline-retest and normative ROC coordinates. Although
the sensitivity of the symptom scores is moderate using this approach, a 1-point difference in scores might not be optimal for making clinical decisions post-concussion. Specifically, concussion-related symptoms are largely dependent on the state of the individual and typically fluctuate over time. This was observed in the current study by the small to medium reliability of symptom scores between baseline and retest in the healthy sample, with the total symptom score having a Pearson correlation coefficient of .41 from baseline to the second time tested (M= 256.8 days), and .32 from baseline to the third time tested (M= 275.8 days), and the symptom severity score having a Pearson correlation coefficient of .30 from baseline to the second time tested, and .22 from baseline to the third time tested. Moreover, a 1-point drop in the SAC could be representative of not knowing the correct date, and similarly, a 1-point drop in the m-BESS is representative of one only one error. As such, the use of reliable change indices or normative comparisons might be more clinically useful than an arbitrary 1-point drop in test scores.

5.2.2 ANAM
In comparison to the ANAM as a combined measure, it was found that the SCAT3 was less sensitive in evaluating concussion in both the acute and post-acute phases of injury, with measures of reaction time yielding maximum sensitivity. This finding is consistent with previous research revealing the utility of the ANAM and particularly the reaction time subtests in evaluating concussion within three days post-injury, however contradicts previous findings of the limited utility of the ANAM if administered 10 days post-concussion (Coldren et al., 2012). As the study conducted by Coldren and colleagues (2012) evaluated the clinical utility of the ANAM in identifying concussion in military blast injuries, the current study provides evidence that the ANAM sports medicine battery is more appropriate and has increased sensitivity in the athletic context. As such, it should be used supplementary to the SCAT3 across the concussion trajectory to detect more subtle deficits and inform clinical management.

5.2.3 Baseline Versus Normative Data
Currently, there is little empirical evidence to support the implementation of baseline testing, however it is thought to add useful information to test interpretation post-concussion (Kontos, Sufrinko, Womble, & Kegel, 2016; McCrory et al., 2013). This research sought to determine whether the utility of the SCAT3 was influenced by the presence or absence of baseline test data.
While although Chapter 3 identified population-specific moderators of baseline SCAT3 scores that could affect post-concussion test interpretation, Chapter 4 demonstrated that the use of baseline data does not necessarily enhance the sensitivity and overall utility of the SCAT3 in identifying concussion compared to normative data. The results of these studies are therefore consistent with current guidelines, providing empirical support against the necessity of baseline testing, however may be helpful in cases where athletes have a medical history, such as a previous diagnosis of a mental health disorder or a LD/ADHD.

It was also revealed that the subtests included in the SCAT3 were differentially affected by the comparison of post-concussion scores against either baseline performance or a normative database. Specifically, it was demonstrated that the sensitivity of the total symptom and symptom severity scores were heightened using a normative approach, however the SAC and m-BESS research optimal sensitivity using baseline data. This finding may be associated with the inherent variability of symptom reporting, warranting the comparison to a more stable normative database, whereas gross cognitive and balance functioning is more consistent over time, lending to the enhanced sensitivity of using an athlete’s own pre-injury baseline scores. Therefore although the utility of the combined SCAT3 was more sensitive using a normative approach, it is important for clinicians to consider such findings in order to make informed decisions on whether to implement baseline testing.

5.3 Future research directions

The current set of studies highlighted important directions for future research. Firstly, although the study established that the SCAT3 is clinically useful in concussion assessment in varsity athletes over time, there is limited research on the utility of the Child SCAT3 in athletes under the age of 13. As research has demonstrated differences in baseline SCAT2 scores in younger athletes in a youth sample (Glaviano, Benson, Goodkin, Broshek, & Saliba, 2014b), the sensitivity and specificity of the SCAT3 components scores independently and combined might differ for the child SCAT3. As such, future research in this domain is warranted.

Moreover, the current study recruited a control sample that was matched to the concussed athletes based on history of concussion, whereas most previous research has estimated the
sensitivity and specificity of concussion assessment tools by comparing scores against healthy controls with no previous concussions. A recent paper inquired on this dilemma of which comparison group would be most ideal, opining that when evaluating a patient with a head injury, it is less clinically applicable to compare performance against completely healthy participants, compared to those with similar brain injuries (Duff, 2012). As this has never been investigated empirically, it would be valuable for future research to evaluate which control group is most suitable to maximize the sensitivity of the SCAT3 and other concussion assessment tools.

Lastly, the current set of studies assessed the clinical utility of the SCAT3 within 3 weeks post-concussion, with almost half of the sample being medically cleared at this time point. However, research demonstrates that nearly 1% of individuals who experience a concussion may be diagnosed with post concussion syndrome, with symptom resolution prolonging for more than 3 months. As such, future research should evaluate the clinical utility of this measure within 3 months post-injury, when injuries might be more complex.

5.4 Clinical recommendations

As the SCAT3 is proposed to be updated in the fall of 2016 there are a few clinical recommendations that can be taken from these studies.

Firstly, through the collection of studies and analyses, this thesis has revealed that the SCAT3 is clinically useful for the evaluation of concussion. Although found to be more sensitive in the acute phase of injury, the SCAT3 also appears useful in distinguishing between healthy and concussed athletes 3 weeks following initial injury. Due to the brevity of the test, the SCAT3 may easily be implemented within the concussion clinic to gain an initial understanding of the signs and symptoms most affected in an athlete post-concussion, which can be followed up by a more comprehensive clinical interview or assessment such as the ANAM. Indeed, although the SAC has been found to be sensitive in identifying concussion when used on the sideline (Barr & McCrea, 2001), the current study revealed that this brief cognitive measure lacks sensitivity if administered more than 3 days post-concussion. As such, a more comprehensive and multidimensional approach is warranted for evaluating concussion across the concussion sequelae.
Similarly, recent research has demonstrated that the total number and severity of symptoms were significantly lower using a clinically guided open-ended interview compared to a clinically guided PCSS in an adolescent sample of concussed athletes (Elbin et al., 2016). As such, the administration of the structured PCSS included in the SCAT3 is paramount to gaining an accurate understanding of the athlete’s concussion to facilitate proper management.

With the refurbishment of the SCAT3 approaching, the current research also highlights some potential limitations of that SCAT3 that could be modified to improve the utility of this measure. Specifically, Chapter 3 revealed that a history of LD or ADHD was associated with reduced baseline SAC performance, however recent research has identified that a LD or ADHD diagnosis differentially affects neurocognitive performance (Elbin et al., 2013). Therefore, future revisions of the SCAT3 should consider inquiring on an athlete’s history of LD or ADHD separately.

Moreover, although previous research has found moderate sensitivity of the original BESS (Riemann & Guskiewicz, 2000), the current study revealed low sensitivity of the modified version. As such, implementation of the original BESS that includes both the hard and foam testing surface should be considered in future revisions to the SCAT3 if feasible on the sideline.

Findings from this study also highlighted the added sensitivity of reaction time tests within the ANAM for detecting more subtle neurocognitive deficits following concussion. As such, reaction time tests should be considered for inclusion in future SCAT3 revisions if feasible to enhance the utility.

Lastly, a large amount of recent research has concentrated on the potential role of vision assessments in facilitating concussion evaluation (Heinmiller & Gunton, 2016). For instance the King-Devick test is a sideline measure used to quickly detect deficits in eye saccadic movements, attention, and language that has been found to correlate with SAC scores post-concussion (King, Gissane, Hume, & Flaws, 2015). Recent research has suggested that the King-Devick test may increase the sensitivity of identifying concussion on the sideline and in detecting subtle changes as time post-injury extends when used in combination with brief tests assessing cognitive and balance functioning (Ventura, Balcer, & Galetta, 2015). As such, amendments to
the SCAT3 should consider amalgamating such vision assessments as the King-Devick test to enhance its utility.
Chapter 6

Conclusions and Implications

The current set of studies established that the SCAT3 is a clinically useful tool to evaluate concussion in the sport context within 3 weeks post-injury, however may be limited past this timeframe.

Although baseline testing has previously been recommended as the mainstay for concussion assessment, the results from this research demonstrate that the sensitivity yielded from the use of normative SCAT3 data is equivalent to when baseline testing is performed. Clinicians in Canada may use the large Canadian normative database developed from this research to inform post-concussion SCAT3 test interpretation when baseline data is lacking.

Importantly, baseline SCAT3 values were moderated by sex, a history of a psychiatric disorder, a family history of mental health, a history of a diagnosed learning disability or ADHD, and years of education.

As such, clinicians must consider the utility of the SCAT3 on an individual basis, while taking into account the test’s psychometric properties and moderating factors when interpreting post-concussion results. It is recommended that the SCAT3 as a multifaceted concussion assessment tool be implemented alongside other comprehensive evaluations in order to guide clinical management and facilitate recovery to maximize brain health and well-being.
References


Millan, M. J., Agid, Y., Brüne, M., Bullmore, E. T., Carter, C. S., Clayton, N. S., … Young, L. J.


Appendices

Appendix 1. Information sheet/consent form

ETHICS REVIEW INFORMATION SHEET/CONSENT FORM

Title of research project: “Recovery Following Sport Concussion”

Investigators:
Michael Hutchison, PhD (Principal Investigator)
Faculty of Kinesiology and Physical Education
University of Toronto
e-mail: michael.hutchison@utoronto.ca
phone: (416) 948-4050
Paul Comper, Ph.D., Toronto Rehab – University Health Network
Lynnae Mainwaring, Ph.D., University of Toronto
Patrick Quaid, MCOptom, Ph.D., Guelph Vision Therapy Centre
Shawn Rhind, Ph.D., Defence Research and Development Canada
Doug Richards, M.D., University of Toronto
Tom Schweitzer, Ph.D., St. Michael’s Hospital

Background & Purpose of Research:
Clinical strategies for the management of sports-related concussion have increased in recent years. Most recently, there has been a trend toward more advanced and individualized approaches to managing concussions.

The purpose of this study is to examine various objective markers in order to determine the natural course of recovery from concussion. The addition of these objective markers to our existing clinical protocol will significantly extend and strengthen our understanding of concussion. In the end, the study will provide insight for the development of evaluation tools that are useful and practical for physicians and allied health professionals.

Eligibility:
To participate in this study you must be participating in sports at the university level that are considered to be high risk for concussion (i.e., Basketball, Field Hockey, Football, Lacrosse, Hockey, Soccer, Rugby, and Volleyball). Both male and female athletes are invited to participate in all aspects of this study.

Procedures:
Participation in this study requires you to complete a collection of tests at the beginning of your competitive season, also known as the “baseline” session. The test battery includes: cognitive testing, vision and balance assessment, a blood sample, and magnetic resonance imaging (MRI), which provides a picture of your brain. If you receive a concussion during your athletic season, you will complete the battery of tests again between 3-7 days post-injury, and once a week until you have received medical clearance to return to play. The last assessment will take place one month after you returned to play. The testing assessment schedule allows us to monitor the unique recovery course for each athlete, since the duration of this process can be highly variable. In total, four MRI assessments will be completed. The first MRI session will take place at baseline, the second will be 3-7 days post-injury, the third will be done at the time of medical clearance, and the fourth will take place one month after return-to-play.

Version 1: August 2014
Appendix 1. contd.

Athletes with musculoskeletal injury (such as sprained ankle or wrist) will follow the same protocol as concussed athletes. Athletes who complete baseline testing and do not sustain an injury during the study period may or may not be recruited as a healthy control.

**Neuropsychological (NP) Testing:**
You will be seated and asked to complete a series of computerized cognitive tests designed to measure the following brain functions: reaction time, attention, memory, visuospatial ability. This test battery is referred to as the Automated Neuropsychological Assessment Metrics (ANAM). It is a self-directed computerized cognitive test battery and you will be instructed to complete the tests as quickly and accurately as possible, which normally takes approximately 25 minutes to complete.

**Vision Assessment:**
Our vision assessment battery includes a variety of tests measuring different eye functions, such as depth perception and visual acuity, as well as self-report questionnaires, which ask about vision difficulties. Many of the tests included are commonly used by eye doctors (optometrists and ophthalmologists) to determine vision function with one or both eyes. The tests are non-invasive and do not present risk of injury or discomfort. Subjects will be assessed with their monocular habitual acuity (i.e. either with or without glasses, depending on what they usually use). This process normally takes about 40 minutes.

**Balance Testing:**
The balance test will assess your ability to maintain postural control during two different conditions: 1) Standing with feet together, eyes open and 2) Standing with feet together, eyes closed. In both conditions you will be asked to stand barefoot in the middle of a force plate (20 x 20 inches, elevated 5 cm off the ground) with the insides of your feet touching and your arms comfortably behind your back. You will begin each trial by staring at a point directly in front of you and in the case of the ‘eyes closed’ condition; you will close your eyes following a “go” command from the researcher. Each trial will be last 60 seconds. There will be a total of 4 completely random trials administered and you will be allowed to rest in between trials if necessary.

**Blood Serum Sample:**
Physician or research staff member with venipuncture (blood-drawing) training will collect a small amount of blood (approximately 40 mL) using a needle and vacutainer tubes. This process normally takes approximately 5 minutes.

**Magnetic Resonance Imaging (MRI):**
Medical imaging will be conducted at St. Michael’s Hospital. The MRI technique uses magnets and radio waves to construct a picture of your brain on a computer.

Prior to the scan, you will be asked to complete a MRI Screening Form for metallic objects which could be affected by the MRI. A medical imaging technician will review this form with you before the scan begins. Participants who may be affected by the MRI will be excluded from the study.

For the procedure, you will be asked to lie on a padded bed that will be moved into a tunnel-like machine for the MRI scan of your brain. Since you will be inside the machine during the scan, and a screen will be in place for viewing the visual images, you may not be able to see the technicians operating the machine or the investigators. However, there is an intercom system that will allow you to talk with them at any time. If you feel uncomfortable during the scan and you wish to discontinue the procedure, you will be taken out of the machine immediately at your request.

We will obtain a series of MRI scans, separated by short breaks, and the entire procedure will take approximately 90 minutes. You should try to remain as still as possible during each scan. Movement will not be dangerous to you in any way, but would blur the picture of your brain. You will hear Version 1: August 2014
Appendix 1. contd.

The collected blood sample will be coded with a Subject ID code which will not identify you and the sample will be stored in the principal investigator’s research space at the University of Toronto. For processing and analyses samples will be sent to Defense Research and Development Canada (DRDC) – Toronto. Personnel at DRDC will not be able to link your blood sample to your identity.

MRI data and image representations will be stored on a secure network server that is password protected in an anonymous database referencing only Subject ID, which is common practice for Dr. Schweizer’s imaging laboratory. The MRI Screening Form, and all other study documents both paper and electronic information will be stored within principal investigator’s research space at the University of Toronto. Data will be stored securely stored in a locked room that is accessible only to investigators and within this room they will be stored in locked cabinets and electronic files on computers will be password protected. Data will be kept for a period of 6 years. In the event that you withdraw from the study, any data collected up to that point will be kept for a period of 6 years.

Following the initial intake, your study records will be made anonymous by the use of a confidential numeric Subject ID. Electronic files will be stored securely on University of Toronto or hospital networks or securely on any portable electronic devices.

Please note that confidentiality can only be guaranteed to the extent permitted by law.

What are the costs of participating in this study?
Participating in this study may result in costs to you for parking, transportation, and meals.

You will be compensated $20 per session for completing the battery of tests at the University of Toronto on a weekly basis until study completion to offset any expenses incurred (such as travel expenses) as a result of taking part in this study. We expect weekly sessions to take between 1.5-2.0 hours.

You will be compensated $50 per MRI assessment at St. Michael’s Hospital to offset any expenses incurred by travel or preparation time as a result of taking part in the MRI portion of this study.

Publication of Research Findings:
Following the conclusion of this study the results may be published. No information that could reveal you as a participant will be disclosed in any publication.

New Findings:
If anything comes to light during the course of this research, which may influence your decision to continue, you will be notified.

Rights of Subjects:
You have no legal rights by participating in this trial. If you have any questions regarding your rights as a participant you may contact:

Office of Research Ethics, Health Sciences
2nd Floor, McMurtrich Building
12 Queen’s Park Cres W
Toronto, Ontario, M5S 1S6
Tel: 416-946-3273

Dissemination of Findings:
The results of this study will be presented or published in collective form. Your personal information will not be used and as such, any published results cannot identify you as an individual. As a research participant you have the right to request a copy of the final report of the findings of this research study.

Version 1: August 2014
Appendix 1. contd.

Title of research project: “Recovery Following Sport Concussion”

By signing this form, I agree that:

- The study has been explained to me.
- All of my questions were answered.
- The possible harms and discomforts and the possible benefits (if any) of this study have been explained to me.
- I understand that I have the right not to participate and the right to stop at any time.
- I understand that I may refuse to participate without any problems.
- I have a choice of not answering any specific questions.
- I am free now, and in the future, to ask any questions about the study.
- I understand that there is a very slight risk while I give blood and that there may be minimal chance of infection. These discomforts are brief and temporary.
- I have been told that my personal records will be kept confidential.
- I have been informed that the hazards of Magnetic Resonance Imaging (MRI) are minimal but a small percentage of individuals may become nervous because of loud noises and the confined space. A MRI screening form will be completed to ensure appropriateness prior to scanning procedure.
- I understand that I will receive a signed copy of this consent form.
- I agree that my data may be used for future testing in similar research projects.

Name (PRINT NAME): __________________________________________

A) I hereby consent to participate in the study.

Signature: ___________________________ Date: ________________

B) I do not consent to participate in the study at this time. However, during the season, if I sustain a concussion or musculoskeletal injury, I give my permission for the sports medicine personnel at the University of Toronto to report the injury to one of the investigators of the research team and contact me then to ask if I would like to participate in the research and post-injury testing at that time.

Signature: ___________________________ Date: ________________

I, the undersigned, have fully explained the study to the above participant.

Investigator/Designate Name: __________________________________

Investigator/Designate Signature: ___________________________ Date: ________________

Study Contact:
If you have any questions about this study now or in the future, or in the case of a study-related injury please call Dr. Michael Hutchison at (416) 946-4050 or michael.hutchison@utoronto.ca

Version 1: August 2014
**Appendix 2  Copy of SCAT3 form with major components: PCSS, SAC, m-BESS**

**BACKGROUND**

<table>
<thead>
<tr>
<th>Name:</th>
<th>Date:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Examin:</td>
<td></td>
</tr>
<tr>
<td>Sport/team/school:</td>
<td>Date/time of injury:</td>
</tr>
<tr>
<td>Age:</td>
<td>Gender: M F</td>
</tr>
<tr>
<td>Years of education completed:</td>
<td></td>
</tr>
<tr>
<td>Dominant hand:</td>
<td>right left neither</td>
</tr>
<tr>
<td>How many concussions do you think you have had in the past?</td>
<td></td>
</tr>
<tr>
<td>When was the most recent concussion?</td>
<td></td>
</tr>
<tr>
<td>How long was your recovery from the most recent concussion?</td>
<td></td>
</tr>
<tr>
<td>Have you ever been hospitalized or had medical imaging done for a head injury?</td>
<td>Y N</td>
</tr>
<tr>
<td>Have you ever been diagnosed with headaches or migraines?</td>
<td>Y N</td>
</tr>
<tr>
<td>Do you have a learning disability, ADHD, ADD/ADHD?</td>
<td>Y N</td>
</tr>
<tr>
<td>Have you ever been diagnosed with depression, anxiety or other psychiatric disorder?</td>
<td>Y N</td>
</tr>
<tr>
<td>Has anyone in your family ever been diagnosed with any of these problems?</td>
<td>Y N</td>
</tr>
<tr>
<td>Are you on any medications? If yes, please list:</td>
<td></td>
</tr>
</tbody>
</table>

**SYMPTOM EVALUATION**

**How do you feel?**

"You should score yourself on the following symptoms, based on how you feel now".

<table>
<thead>
<tr>
<th>Symptom</th>
<th>None</th>
<th>Mild</th>
<th>Moderate</th>
<th>Severe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Headache</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Pressure in head</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Neck Pain</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Nausea or vomiting</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Dizziness</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Blurred vision</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Sensitivity to light</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Sensitivity to noise</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Feeling slowed down</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Feeling like &quot;in a fog&quot;</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Don't feel right</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Difficulty concentrating</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Difficulty remembering</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Fatigue or low energy</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Confusion</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Drowsiness</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Trouble falling asleep</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>More emotional</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Irritability</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Sadness</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Nervous or Anxious</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

**COGNITIVE & PHYSICAL EVALUATION**

**Cognitive assessment**

**Standardized Assessment of Concussion (SAC)**

<table>
<thead>
<tr>
<th>Orientation</th>
<th>Trail 1</th>
<th>Trail 2</th>
<th>Trail 3</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Immediate memory**

<table>
<thead>
<tr>
<th>List</th>
<th>Trial 1</th>
<th>Trial 2</th>
<th>Trial 3</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

**Total Immediate memory score**: 

**Concentration: Digit Backward**

<table>
<thead>
<tr>
<th>List</th>
<th>Trial 1</th>
<th>Trial 2</th>
<th>Trial 3</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

**Neck Examination:**

**Range of motion**

**Tenderness**

**Upper and lower limb sensation & strength**

**Findings:**

**Balance examination**

**Do one or both of the following tests:**

**Footwear** (shoes, barefoot, braces, tape, etc.)

**Modified Balance Error Scoring System (BESS) testing**

**With foot was tested (i.e. which is the non-dominant foot)**

**Testing surface (hard floor, field, etc.)**

**Condition**

**Double leg stance:**

**Errors**

**Single leg stance (non-dominant foot):**

**Errors**

**And/or:**

**Tandem gait**

**Time (best of 4 trials):**

**Coordination examination**

**Upper limb coordination**

**Which arm was tested:**

**Coordination score**

**SAC Delayed Recall**

**Delayed recall score**: 

**Scoring on the SCAT3 should not be used as a stand-alone method to diagnose concussion, measure recovery or make decisions about an athlete's readiness to return to competition after concussion. Since signs and symptoms may evolve over time, it is important to consider repeat evaluation in the acute assessment of concussion.**